Федеральное государственное бюджетное образовательное учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

ДАЛЬНЕВОСТОЧНЫЙ ИНСТИТУТ УПРАВЛЕНИЯ

УТВЕРЖДЕНО
Директор
Меркулов Е.С.

Электронная подпись

ПРОГРАММА МАГИСТРАТУРЫ
Бухгалтерский учет и налогообложение

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ,
реализуемой без применения электронного (онлайн) курса

Б1.О.04 Машинное обучение
(индекс, наименование дисциплины, в соответствии с учебным планом)

38.04.01 Экономика
(код, наименование направления подготовки)

Очная, заочная
(форма(ы) обучения)

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 43B0028FD23306BCF414031E3FCDC222
Владелец: Меркулов Евгений Сергеевич
Действителен: с 29.01.2025 до 24.04.2026

Год набора - 2025

Хабаровск, 2025 г.

Автор-составитель:
кандидат экономических наук, доцент кафедры экономики и цифровых технологий Пынько Л.Е.

Заведующий кафедрой
экономики и цифровых технологий, кандидат экономических наук, доцент
Поливаева О.Г.

РПД «Машинное обучение» одобрена на заседании кафедры экономики и цифровых технологий. Протокол от 21.04.2025 № 12

СОДЕРЖАНИЕ

1 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы……. 4
2 Объем и место дисциплины в структуре образовательной программы…… 4
3 Содержание и структура дисциплины……. 5
4 Материалы текущего контроля успеваемости обучающихся……..... 12
5 Оценочные материалы промежуточной аттестации по дисциплине…… 60
6 Методические материалы для освоения дисциплины…… 64
7 Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет» …… 67
7.1 Основная литература …… 67
7.2 Дополнительная литература…… 68
7.3 Нормативные правовые документы и иная правовая информация …….. 69
7.4 Интернет-ресурсы…… 70
7.5 Иные источники…… 71
8 Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы…… 72

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

1.1 Дисциплина Б1.О.04 «Машинное обучение» обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код компонента компетенции	Наименование компонента компетенции
ОПК ОС-3	Способен применять продвинутые инструментальные методы экономического анализа при решении практических и (или) исследовательских задач	ОПК ОС-3.1	Способен использовать современные инструментальные методы экономического анализа
ОПК ОС-5	Способен использовать современные информационные технологии и программные средства при решении практических и (или) исследовательских задач	ОПК ОС-5.1	Способен использовать современные информационные технологии и программные средства для решения профессиональных задач

1.2 В результате освоения компетенции у студентов должны быть сформированы:

ОТФ/ТФ/трудовые или профессиональные действия	Код компонента компетенции	Результаты обучения
	ОПК ОС-3.1	на уровне знаний : знает современные инструментальные методы экономического анализа для решения исследовательских и практических задач; на уровне умений : использует современные инструментальные методы экономического анализа для решения исследовательских и практических задач
	ОПК ОС-5.1	на уровне знаний : – представление о современных информационных технологиях и программных средств для решения профессиональных задач; на уровне умений : – использует современные информационные технологии и программные средства для решения профессиональных задач.

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины: общая трудоемкость дисциплины 5 з.е. по учебному плану, что составляет 180 ак. ч. / 135 астр.час., из них:

- для очной формы обучения: на контактную работу выделено 64 акад. часов / 48 астр.час. (из них: 32 акад. часов / 24 астр. час. лекционных занятий, 32 акад. часов / 24 астр. час. практических занятий), 80 акад. часов / 60 астр. час – на самостоятельную работу обучающихся, 36 акад. часов / 27 астр. час – на контроль (самостоятельная работа обучающихся в период зачетно-экзаменационной сессии).

- для заочной формы обучения: на контактную работу выделено 12 акад. часов / 9 астр.час. (из них: в рамках установочной сессии - лекционные занятия 4 акад. часов / 3 астр. час., самостоятельная работа студентов 32 ак.ч/24 астр.ч.; в рамках зимней сессии: 2 ак. ч./1,5 астр. ч. лекционных занятий, 6 акад. часов / 4,5 астр. час. практических занятий), 127 акад. часов / 95,25 астр. час – на самостоятельную работу обучающихся, 9 акад. часов / 6,75 астр. час – на контроль (самостоятельная работа обучающихся в период зачетно-экзаменационной сессии).

Место дисциплины в структуре ОП ВО

В соответствии с учебным планом дисциплина Б1.О.04 «Машинное обучение» изучается: на очной форме обучения – на 1 курсе в 1 семестре; на заочной форме обучения – на 1 курсе в установочную и зимнюю сессию.

Форма промежуточной аттестации по дисциплине в соответствии с учебным планом - экзамен.

3 Содержание и структура дисциплины

3.1. Структура дисциплины

Очная форма обучения

№ п/п	Наименование тем (разделов)	Объем дисциплины (модуля), час.						Форма текущего контроля успеваемости*, промежуточной аттестации**
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л/ ДОТ	ЛР/ ДОТ	ПЗ/ ДОТ	КСР		
	Раздел I. Современные подходы организации машинного обучения и нейросетевого программирования	50	8	-	12	-	30	
Тема 1	Возможности средств машинного обучения и средств кластеризации для реализации задач оценки рисков и экономического моделирования решения проблем в профессиональной деятельности	16	2	-	4	-	10	О, ОПЗ
Тема 2	Технические и программные средства реализации машинного обучения, возможные проекты применения машинного обучения для оценки рисков	18	2	-	4	-	12	О, ОПЗ
Тема 3	Основные методы машинного обучения.	16	4	-	4	-	8	О, Т
	Раздел II. Нейросетевые технологии в	48 (2*)	12	-	10 (2*)	-	26	

	практической деятельности сферы экономики и производства средствами машинного обучения							
Тема 4	Нейросетевые технологии анализа, прогнозирования финансовых показателей и оценки финансового риска средствами машинного обучения	16	4	--	4	-	8	О, ОПЗ
Тема 5	Средства интернет-технологий структурирования и организации данных	18 (2*)	4		6 (2*)		8	О, ОПЗ
Тема 6	Системы машинного обучения и нейросетевого анализа в отдельных секторах экономики и производства. Методы машинного обучения при поиске информации в современных системах с помощью систем машинного обучения	14	4	-	-	-	10	О, Т
	Раздел III. Средства машинного обучения в обработке информации производственно-исследовательской и научно-исследовательской деятельности	46 (2*)	12 (2*)	-	10	-	24	
Тема 7	Распределенная обработка информации. Организация интернет- интернет-систем научных и производственных программ	12 (2*)	4 (2*)	-	-	-	8	О, ОПЗ
Тема 8	Применение средств сетевых информационных технологий телекоммуникаций в науке с интеграцией систем машинного обучения для анализа информации	18	4	-	6	-	8	О, ОПЗ
Тема 9	Применение средств сетевых интернет-технологий телекоммуникаций в профессиональном образовании отрасли сферы экономики и производства с интеграцией систем машинного обучения для анализа информации	16	4	-	4	-	8	О, Т
Промежуточная аттестация		**36**	**-**	**-**	**-**	**36**	**-**	Экзамен
Всего:		**180**	**32**	**-**	**32**	**36**	**80**	

Используемые сокращения:

Л - занятия лекционного типа (лекции и иные учебные занятия, предусматривающие

преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся);
ЛР - лабораторные работы (вид занятий семинарского типа;
ПЗ - практические занятия (виды занятий семинарского типа за исключением лабораторных работ);
КСР - индивидуальная работа обучающихся с педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях (в том числе индивидуальные консультации);
ДОТ -занятия, проводимые с применением дистанционных образовательных технологий, в том числе с применением виртуальных аналогов профессиональной деятельности.
СРО - самостоятельная работа, осуществляемая без участия педагогических работников организации и (или) лиц, привлекаемых организацией к реализации образовательных программ на иных условиях.
** формы текущего контроля успеваемости: опрос (О), письменный отчет по практическому занятию (ОПЗ), тестирование (Т)*
*** форма промежуточной аттестации: экзамен*

Заочная форма обучения

№ п/п	Наименование тем (разделов)	Объем дисциплины (модуля), час.						Форма текущего контроля успеваемости*, промежуточной аттестации**
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л/ ДОТ	ЛР/ ДОТ	ПЗ/ ДОТ	КСР		
	Раздел I. Современные подходы организации машинного обучения и нейросетевого программирования	54	2	-	-	-	52	
Тема 1	Возможности средств машинного обучения и средств кластеризации для реализации задач оценки рисков и экономического моделирования решения проблем в профессиональной деятельности	16	2	-	-	-	14	О, ОПЗ
Тема 2	Технические и программные средства реализации машинного обучения, возможные проекты применения машинного обучения для оценки рисков	20	-	-	-	-	20	О, ОПЗ
Тема 3	Основные методы машинного обучения.	18	-	-	-	-	18	О, Т
	Раздел II. Нейросетевые технологии в практической деятельности сферы экономики и производства средствами машинного обучения	54 (2*)	2	-	2 (2*)	-	50	
Тема 4	Нейросетевые технологии анализа, прогнозирования финансовых показателей и оценки финансового риска	16	2	--	-	-	14	О, ОПЗ

	средствами машинного обучения							
Тема 5	Средства интернет-технологий структурирования и организации данных	20 (2*)			2 (2*)		18	О, ОПЗ
Тема 6	Системы машинного обучения и нейросетевого анализа в отдельных секторах экономики и производства. Методы машинного обучения при поиске информации в современных системах с помощью систем машинного обучения	18	-	-	-	-	18	О, Т
	Раздел III. Средства машинного обучения в обработке информации производственно-исследовательской и научно-исследовательской деятельности	63 (2*)	2 (2*)	-	4	-	57	
Тема 7	Распределенная обработка информации. Организация интернет- интернет-систем научных и производственных программ	20 (2*)	2 (2*)	-	-	-	18	О, ОПЗ
Тема 8	Применение средств сетевых информационных технологий телекоммуникаций в науке с интеграцией систем машинного обучения для анализа информации	22	-	-	2	-	20	О, ОПЗ
Тема 9	Применение средств сетевых интернет-технологий телекоммуникаций в профессиональном образовании отрасли сферы экономики и производства с интеграцией систем машинного обучения для анализа информации	21	-	-	2	-	19	О, Т
Промежуточная аттестация		**9**	**-**	**-**	**-**	**9**	**-**	Экзамен
Всего:		**180**	**6**	**-**	**6**	**9**	**159**	

Используемые сокращения:
Л - занятия лекционного типа (лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся);
ЛР - лабораторные работы (вид занятий семинарского типа;
ПЗ - практические занятия (виды занятий семинарского типа за исключением лабораторных работ);
КСР - индивидуальная работа обучающихся с педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях (в том числе индивидуальные консультации);
ДОТ -занятия, проводимые с применением дистанционных образовательных технологий, в

том числе с применением виртуальных аналогов профессиональной деятельности.
СРО - самостоятельная работа, осуществляемая без участия педагогических работников организации и (или) лиц, привлекаемых организацией к реализации образовательных программ на иных условиях.
** формы текущего контроля успеваемости: опрос (О), письменный отчет по практическому занятию (ОПЗ), тестирование (Т)*
*** форма промежуточной аттестации: экзамен*

3.2. Содержание дисциплины

№ п/п темы	Наименование тем (разделов)	Содержание тем (разделов)
	Раздел I. Инструментальные средства компьютерных технологий информационного обслуживания научной и организационной деятельности	
Тема 1	Возможности средств машинного обучения и средств кластеризации для реализации задач оценки рисков и экономического моделирования решения проблем в профессиональной деятельности	Программные сервисы и Системы машинного обучения (СМО, англ. Machine learning, ML). Подходы в формировании прогнозов и автоматического принятия деловых решений средствами нейросетевого программирования и машинного обучения. Искусственный интеллект. Глубокое обучение. Виды нейросетей. Машинное обучение (МО), как тип алгоритма или метода программирования, который позволяет программным системам и бизнес-приложениям стать более эффективными и точными в прогнозировании результатов. Основные процессы, связанные с машинным обучением, напрямую связаны с процессами искусственного интеллекта (ИИ, AI) и интеллектуального анализа данных (ИАД). Системы машинного обучения (СМО), как прикладные приложения, где алгоритмы обучения используются в системе для обеспечения автоматизированного выполнения интеллектуальных бизнес- или производственных задач. Источникам данных для организации машинного обучения Среди систем машинного обучения. Класс систем - Системы глубокого обучения (англ. Deep learning systems).
Тема 2	Технические и программные средства реализации машинного обучения, возможные проекты применения машинного обучения для оценки рисков, возможные проекты применения машинного обучения для оценки рисков	Проект прогнозирования цен на акции с использованием методов усреднения и линейной регрессии. Оценка кредитного риска методами машинного обучения. Анализ и прогнозирование временных рядов цен на акции отечественных компаний. Прогнозирование удовлетворенности клиентов методами машинного обучения. Анализ рынка акций с использованием простых методов машинного обучения. Проект машинного обучения для классификации мошеннических и не мошеннических транзакций. Проект по выявлению мошенничества с кредитными картами методами машинного обучения. Прогнозирование платежеспособности небанковского клиента методами машинного обучения. Проект прогнозирования потребительской ценности методами машинного обучения. Сегментация клиентов методами машинного обучения. Прогнозирование спроса на продукцию методами машинного обучения. Прогнозирование банкротства компании методами машинного обучения. Прогнозирование оттока клиентов методами машинного обучения. Прогнозирование дефолта по кредитной карте методами машинного обучения. Основы KNIME. Среда конструирования нейронных сетей, организация машинного обучения инструментами KNIME.
Тема 3	Основные методы машинного обучения	Виды задач в ML. Два основных подхода в машинном обучении. Обучение с учителем (supervised learning): Классификация (classification). Регрессия (regression). Ранжирование (learning to rank). Системы машинного обучения с учителем. Обучение без учителя (unsupervised learning). Кластеризация (clustering). Уменьшение размерности (dimensionality reduction). Системы машинного обучения без учителя.
	Раздел II. Нейросетевые технологии в практической деятельности сферы экономики и производства средствами машинного обучения	

№ п/п темы	Наименование тем (разделов)	Содержание тем (разделов)
Тема 4	Нейросетевые технологии анализа, прогнозирования финансовых показателей и оценки финансового риска средствами машинного обучения	Один из подходов к анализу и управлению рисками, основанный на нейросетевых технологиях. Необходимость разработки различных методов прогнозирования рисков в условиях цифровой экономики, важность эффективного риск-менеджмента для современного предприятия, использование нейросетевого метода, как эффективного инструмента анализа. Классификаций рисков при построении нейросетевой модели анализа, области прогнозирования рисков, с применение нейронных сетей, традиционные методы анализа рисков, особое внимание уделено эконометрическим методам.
Тема 5	Средства компьютерных технологий структурирования и организации данных	Обработка числовой информации. Методология работы с числовой информацией. Электронные таблицы. Возможности, принципы и основные приемы работы с электронными таблицами. Технология обработки информации на основе табличных процессоров. Способы систематизации, хранения, обработки и представления числовой информации с помощью электронных таблиц. Обеспечение качественного анализа данных, поиска закономерностей и выработки правильного и оперативного решения через ведение электронных журналов и ведомостей, построение графиков и диаграмм. Технология баз данных. Базы данных (БД). Примеры баз данных учебно-методического назначения. Методы поиска информации в базах данных с помощью систем машинного обучения. Интегрированные программные пакеты. Компьютерные технологии обработки образовательной информации на основе табличных процессоров, использования систем управления базами данных, интегрированных программных пакетов. Особенности современных технологий решения задач текстовой и графической обработки, табличной и математической обработки, накопления и хранения данных. Методы поиска информации в современных системах с помощью систем машинного обучения. Представление знаний. Базы знаний (БЗ). Экспертные системы (ЭС). Примеры экспертных систем соответствующей научной области. Интеллектуальные информационные системы (ИИС).
Тема 6	Системы машинного обучения и нейросетевого анализа в отдельных секторах экономики и производства. Методы машинного обучения при поиске информации в современных системах с помощью систем машинного обучения.	*Росатом* В Росатоме разработали цифровую систему управления атомной отраслью "Навигатор". Она работает на основе технологий с применением искусственного интеллекта. *ММК-Метиз* На ММК-МЕТИЗ испытывают беспилотный «Урал». В течение всего этого времени шла подготовка беспилотника к работе на промышленной площадке: выстраивался маршрут передвижения между производствами и цехами завода, автономная парковка на местах разгрузки, объезд препятствий на пути следования. *Газпромнефть* "Газпромнефть - смазочные материалы" - оператор бизнеса масел "Газпром нефти" - запустил первую в России цифровую систему создания многокомпонентных рецептур моторных масел с помощью искусственного интеллекта. «Газпром нефть» начала перевозки грузов в Арктике беспилотным автотранспортом. *Северсталь* «Северсталь» внедрила программный комплекс по управлению темпом прокатки и выдачей слябов из печей на основе моделей машинного обучения. *ОДК Ростех* "Объединенная двигателестроительная корпорация" "Ростеха" внедрила новый способ люминесцентного контроля качества деталей авиационных двигателей с применением машинного зрения и нейросетевых технологий. *Норникель* На "Норильской обогатительной фабрике" заработал умный помощник флотаторов. Интеллектуальная система внедрена на медистой цепочке "Норильской обогатительной фабрики" (НОФ).

№ п/п темы	Наименование тем (разделов)	Содержание тем (разделов)
		- Быстринский комбинат установил систему контроля за использованием средств индивидуальной защиты, чтобы повысить безопасность на производстве. *Северсталь-метиз* На череповецком заводе компании "Северсталь-метиз" применили технологию искусственного интеллекта. Ее использовали при создании модернизированной системы навигации AGV (Automatic Guided Vehicle – автоматически управляемый транспорт). *ММК* ММК тестирует беспилотные "Камазы". На промплощадке "Магнитогорского металлургического комбината" начались испытания автомобиля "Камаз", оснащённого системой автономного управления. *Роснефть* Специалисты «РН-КрасноярскНИПИнефть» (входит в научно-проектный блок «Роснефти») на основе нейронной сети разработали программный алгоритм, который в 10 раз сокращает время одного из начальных этапов обработки сейсмических данных. Онлайновые библиотечные каталоги: ИНИОН; РГБ; ВГБИЛ; Юридической Научной Библиотеки; университетских библиотек. Книжные интернет-магазины как средство поиска научной информации. Электронные информационные ресурсы: классификация, производители, общие правила работы. Полнотекстовые базы данных мировых агрегаторов научной информации (EBSCO HostWeb, ProQuest, InfoTrac OneFile). Полнотекстовые журнальные базы данных ведущих академических издателей (SAGE Journals Online, Wiley InterScience, Springer Link, Taylor & Francis). Электронная библиотека российских научных журналов eLIBRARY.ru. Электронный архив журнальных публикаций JSTOR. Российская газетная и журнальная периодика в онлайне (базы данных EastView и Интегрум). Полнотекстовые диссертационные базы данных (ProQuest Dissertations & Theses). Онлайновые справочные ресурсы (Oxford Reference Online, Рубрикон). Аналитико-библиографическая база данных Scopus. Научное цитирование. Организация сбора и хранения библиографических данных с помощью специального программного обеспечения: EndNote, ProCite, Reference Manager. Управление научной работой с помощью электронных «напоминаний». Средства идентификации научных статей и книг в электронных информационных ресурсах: DOI; LexisNexis; All England Law Reports – база судебных решений во всех областях права. Базы данных: Iternational Tax Law Reports; Trust Offshore Database; TotalPatent; LexisNexis TU Tracker. Методы поиска информации в современных системах с помощью систем машинного обучения.
	Раздел III. Средства машинного обучения в обработке информации производственно-исследовательской и научно-исследовательской деятельности	
Тема 7	Распределенная обработка информации. Организация компьютерных информационных систем научных и производственных программ	Информационные технологии в научной деятельности. Автоматизация эксперимента, статистической обработки данных, подготовки научных публикаций. Организация проектной деятельности аспирантов в сетях. Методика отбора тематики, организация групп аспирантов в сетях. Информационные технологии образовательных программ. Технология применения электронных образовательных программ. Информационные технологии документационного обеспечения образовательных программ. Образовательные электронные издания и ресурсы. Подготовка дидактических и методических материалов средствами пакета программ Microsoft Office. Подготовка публикаций с помощью настольной издательской системы Microsoft Publisher. Организация учебной проектной деятельности студентов с использованием КТ.

№ п/п темы	Наименование тем (разделов)	Содержание тем (разделов)
		Проблемы технологий в учебном процессе. Теоретико-методологические основы технологизации процесса обучения. Образовательные и обучающие технологии на современном этапе. Проблемы и перспективы информатизации высшего образования. Компьютерные системы в деятельности палат Федерального Собрания Российской Федерации. Автоматизированная система обеспечения законодательной деятельности, ее назначение и функции. Электронные картотеки как интегрированный информационный ресурс Российской Федерации. Стратегические ориентиры и актуальные задачи развития электронного правительства в России. Обучение с подкреплением. Глубокое обучение.
Тема 8	Применение средств сетевых информационных технологий телекоммуникаций в науке с интеграцией систем машинного обучения для анализа информации	Телекоммуникации. Компьютерные телекоммуникации. Локальные и глобальные компьютерные сети. Распределенная обработка информации. Организация компьютерных информационных систем научных и образовательных программ. Обработка данных в сетях. Возможности систем машинного обучения для обработки данных в сетях. Сетевые информационные технологии. Виды компьютерных сетей. Отраслевые и профессионально-ориентированные сети. Почтовые службы. Адресация электронных сообщений. Пользовательский интерфейс почтовых клиентских приложений и настройка их основных параметров. Принципы организации глобальных сетей. Специальные методы поиска информации в мировой глобальной сети Интернет. Возможности систем машинного обучения для поиска в Интернет открытой финансовой информации и данных, представляющих важность в анализе финансовых рисков. Размеченные данные. Частично контролируемое обучение. Алгоритм опорных векторов. Метод опорных векторов. Метод классификатора с максимальным зазором. Модель обнаружения банковского мошенничества. Эвристические алгоритмы. Модель отказа в выдаче кредита.
Тема 9	Применение средств сетевых интернет-технологий телекоммуникаций в профессиональном образовании отрасли сферы экономики и производства с интеграцией систем машинного обучения для анализа информации	Метод опорных векторов. Метод классификатора с максимальным зазором. Модель обнаружения банковского мошенничества. Эвристические алгоритмы. Модель отказа в выдаче кредита. Дистанционное обучение (ДО) и открытое образование. Дистанционные образовательные технологии (ДОТ). Основные принципы дистанционного обучения: гибкость в выборе места и времени обучения, возможность обучения без отрыва от основной деятельности, доступность для живущих в отдаленной местности, свобода выбора учебного заведения и дисциплин для обучения. Решение проблем обеспечения возможности для получения высшего образования и обучения на протяжении всей жизни, предоставления учащимся права свободного выбора места, времени и технологий обучения в рамках системы открытого образования. Возможности систем машинного обучения для организации информации и ее анализа в ДО. Особенности применения компьютерных телекоммуникаций в образовании. Образовательные услуги сети Интернет. Ресурсы Интернет. Дидактические свойства Интернет. Дидактические функции компьютерных телекоммуникаций. Организация дистанционного обучения с использованием программной оболочки. Разработка электронных учебно-методических комплексов. Технологии компьютерного тестирования, обработки и интерпретации результатов тестов. Возможности систем машинного обучения для организации учебного тестирования и проверки результатов.

4. Материалы текущего контроля успеваемости обучающихся

4.1. В ходе реализации дисциплины Б1.О.04 «Машинное обучение» используются следующие методы **текущего контроля** успеваемости обучающихся:

Тема №	Тема (раздел)	Методы текущего контроля успеваемости
Раздел I. Современные подходы организации машинного обучения и нейросетевого программирования		
Тема 1	Возможности средств машинного обучения и средств кластеризации для реализации задач оценки рисков и экономического моделирования решения проблем в профессиональной деятельности	О, ОПЗ
Тема 2	Технические и программные средства реализации машинного обучения, возможные проекты применения машинного обучения для оценки рисков	О, ОПЗ
Тема 3	Основные методы машинного обучения.	О, Т
Раздел II. Нейросетевые технологии в практической деятельности сферы экономики и производства средствами машинного обучения		
Тема 4	Нейросетевые технологии анализа, прогнозирования финансовых показателей и оценки финансового риска средствами машинного обучения	О, ОПЗ
Тема 5	Средства интернет- технологий структурирования и организации данных	О, ОПЗ
Тема 6	Системы машинного обучения и нейросетевого анализа в отдельных секторах экономики и производства. Методы машинного обучения при поиске информации в современных системах с помощью систем машинного обучения.	О, Т
Раздел III. Средства машинного обучения в обработке информации производственно-исследовательской и научно-исследовательской деятельности		
Тема 7	Распределенная обработка информации. Организация интернет- интернет- систем научных и производственных программ	О, ОПЗ
Тема 8	Применение средств сетевых информационных технологий телекоммуникаций в науке с интеграцией систем машинного обучения для анализа информации	О, ОПЗ
Тема 9	Применение средств сетевых интернет- технологий телекоммуникаций в профессиональном образовании отрасли сферы экономики и производства с интеграцией систем машинного обучения для анализа информации	О, Т
Промежуточная аттестация		Экзамен (в случае дистанционной формы проведения: тестирование с прокторингом)

** формы текущего контроля успеваемости: опрос (О), письменный отчет по практическому занятию (ОПЗ), тестирование (Т)*

4.2 Типовые материалы текущего контроля успеваемости обучающихся

Типовые оценочные материалы по Теме 1. «Возможности средств машинного обучения и средств кластеризации для реализации задач оценки рисков и экономического моделирования решения проблем в профессиональной деятельности» и Теме 2 «Технические и программные средства реализации машинного обучения, возможные проекты применения машинного обучения для оценки рисков»

Ключевые понятия: Содержание понятий «нейронные сети» и «Нейросетевое прогнозирование» на примере реализации модели производственной функции Кобба-Дугласа и модели Солоу в части инструментов моделирования динамики экономической системы. Реализация инструментов в кросс – платформенном программном пакете Gretl.

Цель занятия №1: Получение практических навыков по анализу направлений развития цифровых технологий; закрепление практических навыков сбора и обработки статистической информации ключевых экономических показателей регионального масштаба, активизация и усвоение знаний построения производственной функции Кобба-Дугласа, прогнозирование на основе модели Солоу валового внутреннего продукта региона с помощью статистического кросс – платформенного программного пакета Gretl.

Ключевые слова и выражения: производственная функция Кобба-Дугласа, модель Солоу, ресурсы, валовый внутренний продукт, труд, капитал, технический прогресс, эффективность капитала и труда, фондовооружённость труда, производительность труда, отдача капитала, среднедушевое потребление, предельная норма замещения, модель Солоу, средняя, предельная и маржинальная производительность, частная эластичность, эластичность производства, технологическая норма замены, модель экономической динамики, динамика фондов накопления и потребления, политика нормы накопления.

Задание. Используя ресурс www.fedstat.ru ЕМИСС составить динамические ряды в ППП Excel, по валовому региональному продукту в основных ценах, оплате труда наемных работников по регионам Российской Федерации, инвестициям в основной капитал (любой регион ДВФО на выбор) с 2005 по 2025 гг. Требуется проанализировать работу хозяйственной системы и выполнить прогноз по основным показателям ее деятельности. С этой целью необходимо:

- определить параметры и сформировать производственную функцию Кобба-Дугласа с учетом технического прогресса;
- рассчитать ВВП на основе модели производственной функции, сравнить фактические данные ВВП с данными, полученными по модели, результаты сравнения оформить таблицей и графиками;
- рассчитать ВВП на основе модели производственной функции при условии отсутствия технического прогресса, сравнить фактические данные ВВП с данными, полученными по модели при отсутствии технического прогресса, результаты оформить таблицей и графиками;
- провести оценку основных характеристик производственной функции – эффективность капитала и труда, предельной нормы замещения;
- построить модель экономической динамики, взяв за основу модель Солоу с линейным и экспоненциальным изменением нормы накопления, рассчитать основные и дополнительные показатели модели;
- изобразить графики динамики основных и дополнительных показателей, полученных в результате проведенных расчетов.

Содержание отчета по работе.

- постановка задачи;
- параметры и производственная функция Кобба-Дугласа с учетом технического прогресса;
- расчет ВВП на основе модели производственной функции, графические представления фактических данных ВВП и данных, полученными по модели;
- расчет ВВП на основе модели производственной функции при условии отсутствия

технического прогресса, графические представления фактических данных ВВП и данных, полученных по модели при отсутствии технического прогресса;

- оценка основных характеристик производственной функции – эффективность капитала и труда, предельной нормы замещения;
- модель экономической динамики Солоу с линейным и экспоненциальным изменением нормы накопления, расчет основных и дополнительных показателей модели;
- графики динамик основных и дополнительных показателей, полученных в результате проведенных расчетов,
- выводы по работе.

Методические указания к выполнению практических работ сформулированы в тексте заданий, выдаваемых обучающимся на занятии.

Ход работы:

Пакет программ GRETL – статистический инструмент для практической реализации сложных вычислительных процедур эконометрического моделирования. В 2002 году пакет GRETL сделан его разработчиком общедоступным, бесплатным продуктом с возможностью дальнейшей доработки открытых кодов. Возможности программы GRETL таковы:

1. основные описательные статистики (среднее арифметическое, медиана, минимальное и максимальное значения, среднеквадратическое отклонение, коэффициент изменчивости (вариации), коэффициент асимметрии, коэффициент эксцесса);
2. проверка нормальности распределения, распределение частот случайной величины, распределение плотности вероятностей, определение коэффициентов корреляции и т.д.;
3. содержит встроенные статистические таблицы для нормального распределения, t-распределения Стьюдента, F-распределения Фишера, хи-квадрат, Пуассона, биномиального и распределения Дарбина – Уотсона. Существует возможность вычисления критических значений, p-value;
4. анализ временных рядов (набор методов оценивания обобщённым Методом наименьших квадратов, модели ARMAX и GARCH и др.;
5. построение линии тренда, коррелограммы, периодограммы; проверка единичных корней, моделирование типа ARIMA, а также процедуры десезонализации X-12-ARIMA и TRAMO);
5. регрессионный анализ (одношаговый метод наименьших квадратов, взвешенный МНКи прочее);
6. метод главных компонент;
7. экспорт и импорт GRETL –Microsoft Excel и текстовые редакторы;
8. построение графиков и др.

Рассмотрим применение ППП GRETL для вычисления параметров производственной функции Кобба – Дугласа при моделировании динамики экономической системы.

Задача. Пусть известны динамические ряды по основному капиталу, численности рабочих и валовому внутреннему продукту (ВВП) за 20-летний период деятельности некоторой хозяйственной системы (данные приведены в таблице 1). Постоянные параметры, характеризующие норму накопления, равны $s_0 = 0{,}2$ и $h = 0{,}02$ соответственно.

Таблица 1 – Динамические ряды по основному капиталу, численности рабочих и ВВП.

Годы	Капитал	Труд	ВВП
1991	670	2497	7357
1992	679	2500	7544
1993	686	2507	7712
1994	694	2515	7901

Годы	Капитал	Труд	ВВП
1995	701	2520	8079
1996	710	2529	8282
1997	715	2532	8446
1998	715	2534	8577
1999	724	2543	8790
2000	732	2547	8989
2001	738	2554	9182
2002	739	2563	9331
2003	740	2571	9482
2004	747	2580	9698
2005	747	2583	9845
2006	750	2592	10027
2007	758	2599	10256
2008	764	2600	10466
2009	773	2605	10715
2010	777	2605	10912

Требуется проанализировать работу хозяйственной системы и выполнить прогноз по основным показателям ее деятельности. С этой целью необходимо:

- определить параметры и сформировать производственную функцию Кобба – Дугласа с учетом технического прогресса;
- рассчитать ВВП на основе модели производственной функции, сравнить фактические данные ВВП с данными, полученными по модели, результаты сравнения оформить таблицей и графиками;
- рассчитать ВВП на основе модели производственной функции при условии отсутствия технического прогресса, сравнить фактические данные ВВП с данными, полученными по модели при отсутствии технического прогресса, результаты оформить таблицей и графиками;
- провести оценку основных характеристик производственной функции – эффективность капитала и труда, предельной нормы замещения;
- построить модель экономической динамики, взяв за основу модель Солоу с линейным и экспоненциальным изменением нормы накопления, рассчитать основные и дополнительные показатели модели;
- изобразить графики динамики основных и дополнительных показателей, полученных в результате проведенных расчетов.

Решение.

1. Определение параметров и формирование производственной функции. Данные из таблицы 8 поместим на лист Excel в блок ячеек A2:D21 (таблица 2). Сохраним файл в формате Excel, например, на рабочем столе.

Таблица 2 – Заполнение данных в MS EXCEL для импорта в GRETL.

	A	B	C	D
1	*t*	*K*	*L*	*Y*
2	1991	670	2497	7357
3	1992	679	2500	7544
4	1993	686	2507	7712
5	1994	694	2515	7901
6	1995	701	2520	8079
7	1996	710	2529	8282
8	1997	715	2532	8446

	A	B	C	D
9	1998	715	2534	8577
10	1999	724	2543	8790
11	2000	732	2547	8989
12	2001	738	2554	9182
13	2002	739	2563	9331
14	2003	740	2571	9482
15	2004	747	2580	9698
16	2005	747	2583	9845
17	2006	750	2592	10027
18	2007	758	2599	10256
19	2008	764	2600	10466
20	2009	773	2605	10715
21	2010	777	2605	10912

Импортируем сохранённый файл в ППП GRETL. Вход в ППП GRETL осуществляется на нажатию на ярлык  gretl. Проделываем цепочку команд *Файл* → *Открыть* → *Пользовательские* (рисунок 1). В диалоговом окне *Gretl: open file* выбираем сохранённый ранее файл с исходными данными для анализа (рисунок 2) → *Открыть*.

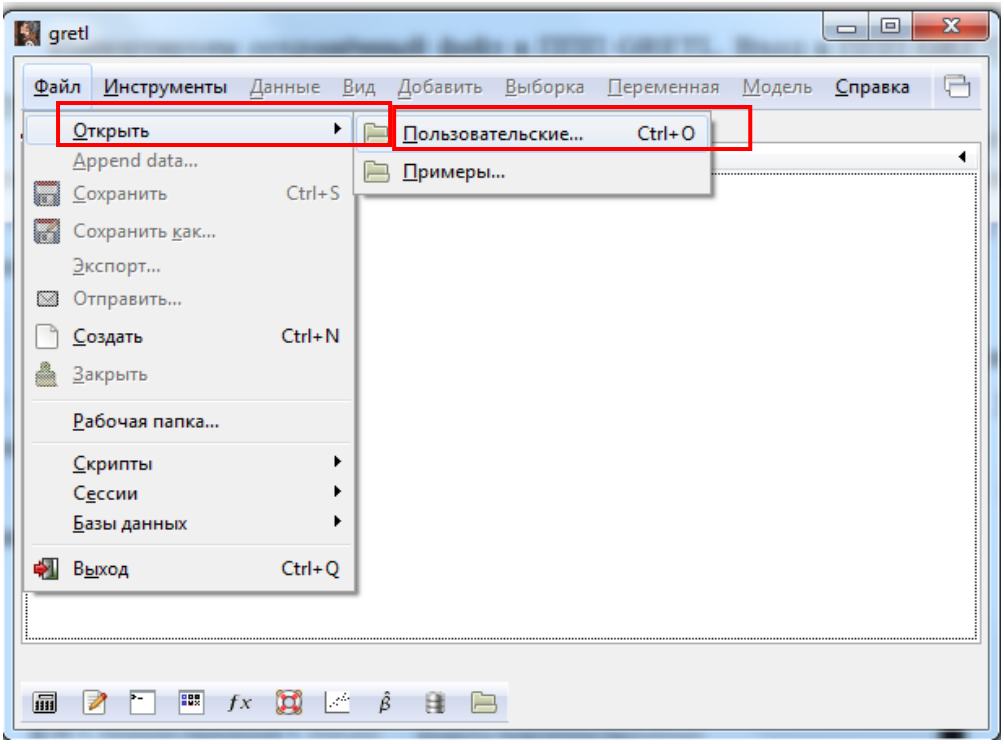


Рисунок 1 – Импорт данных в ППП GRETL

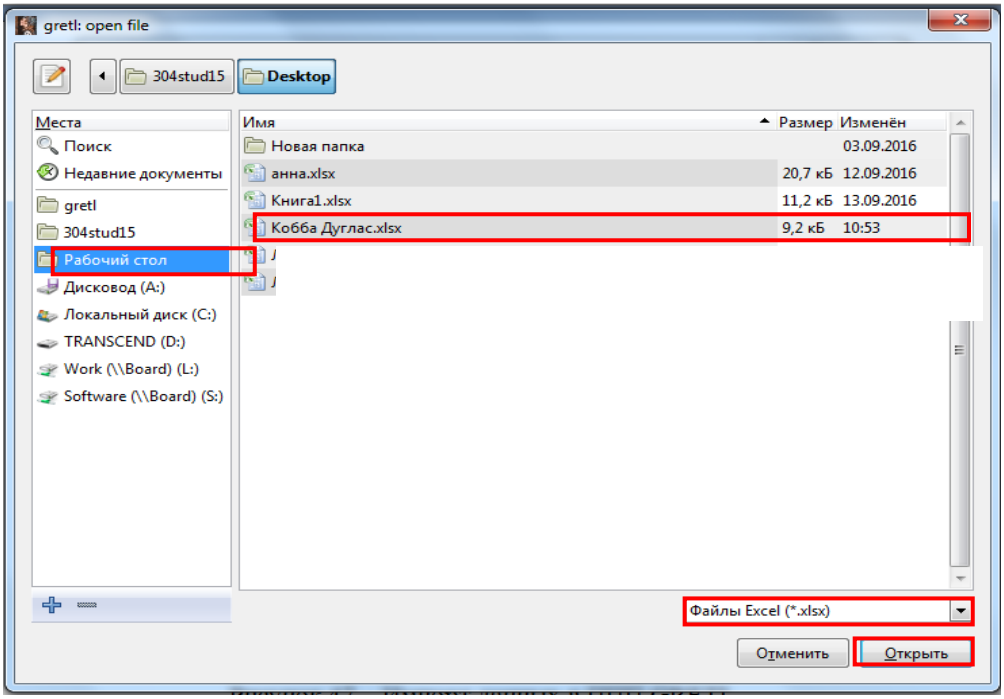


Рисунок 2 – Импорт данных в ППП GRETL из MS EXCEL

Рисунки 3 – 7 демонстрируют импорт рабочей книги формата MS EXCEL в ППП GRETL.

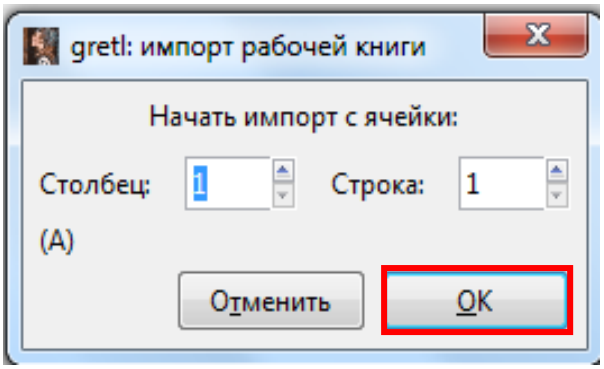


Рисунок 3 – Диалоговое окно *gretl: импорт рабочей книги* в ППП GRETL

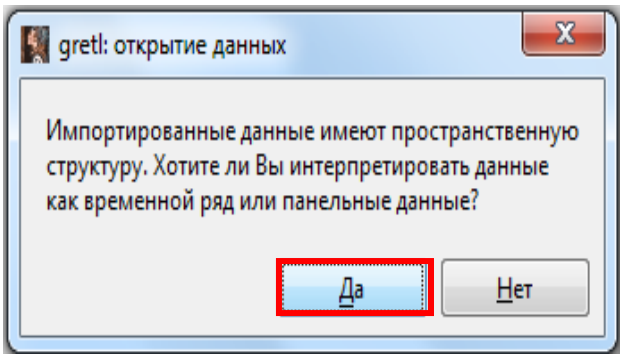


Рисунок 4 – Диалоговое окно *gretl: открытие данных* в ППП GRETL

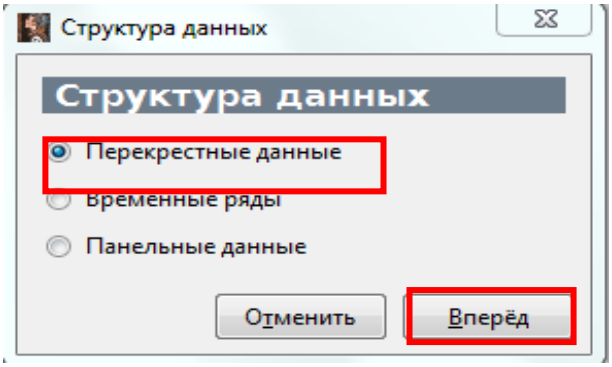


Рисунок 5 – Диалоговое окно *Структура данных* в ППП GRETL

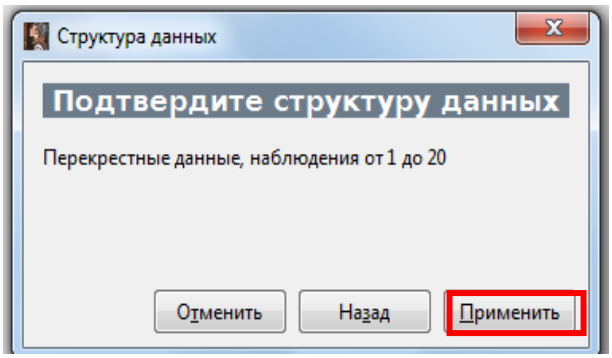


Рисунок 6 – Диалоговое окно *Структура данных* в ППП GRETL

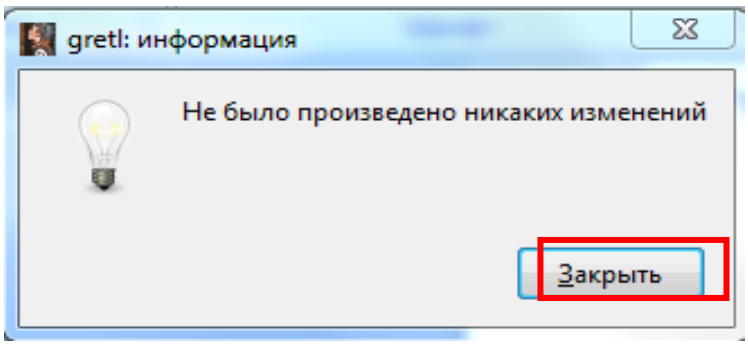


Рисунок 7 – Диалоговое окно *gretl:информация* в ППП GRETL

Производственную функцию Кобба – Дугласа с учетом технического прогресса (**модель Солоу**) будем искать в виде уравнения: $Y = A \cdot K^{\alpha} \cdot L^{\beta} \cdot e^{\gamma(t-t_0)}$ с неизвестными параметрами A, α, β, γ. Логарифмируя эту функцию, получим: $\ln Y = \ln A + \alpha \ln K + \beta \ln L + \gamma(t - t_0)$.

Из этого равенства следует, что значения функции $\ln Y$ линейно зависят от значений $\ln K$, $\ln L$ и $t - t_0$. Поэтому коэффициенты $\ln A$, α, β и γ уравнения можно найти с помощью метода наименьших квадратов.

Для нахождения уравнения $\ln Y = \ln A + \alpha \ln K + \beta \ln L + \gamma(t - t_0)$ в пакете GRETL выполняем цепочку действий.

Добавим переменную временного тренда (рисунок 8).

.

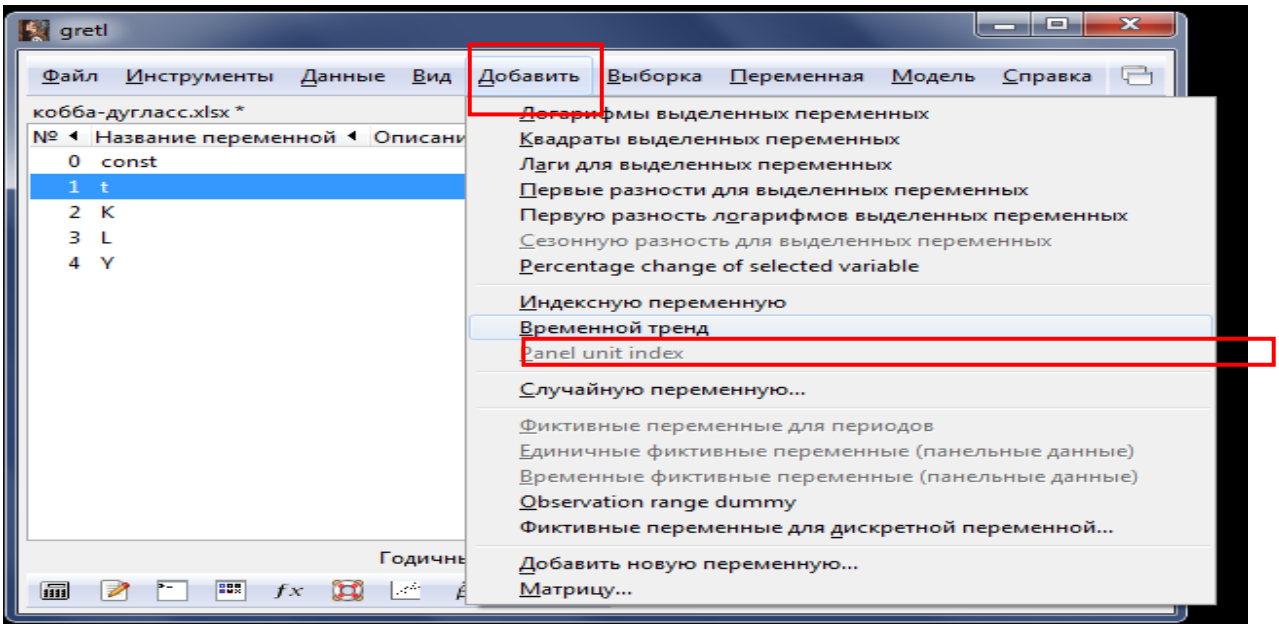


Рисунок 8 – Добавление переменной линейного тренда в ППП GRETL

Щёлкаем правой кнопкой мыши в окне переменных, выбираем *Добавить новую переменную* (рисунок 9). Добавляем новую переменную, заполняем диалоговое окно согласно рисунку 10 ⟶ *OK*.

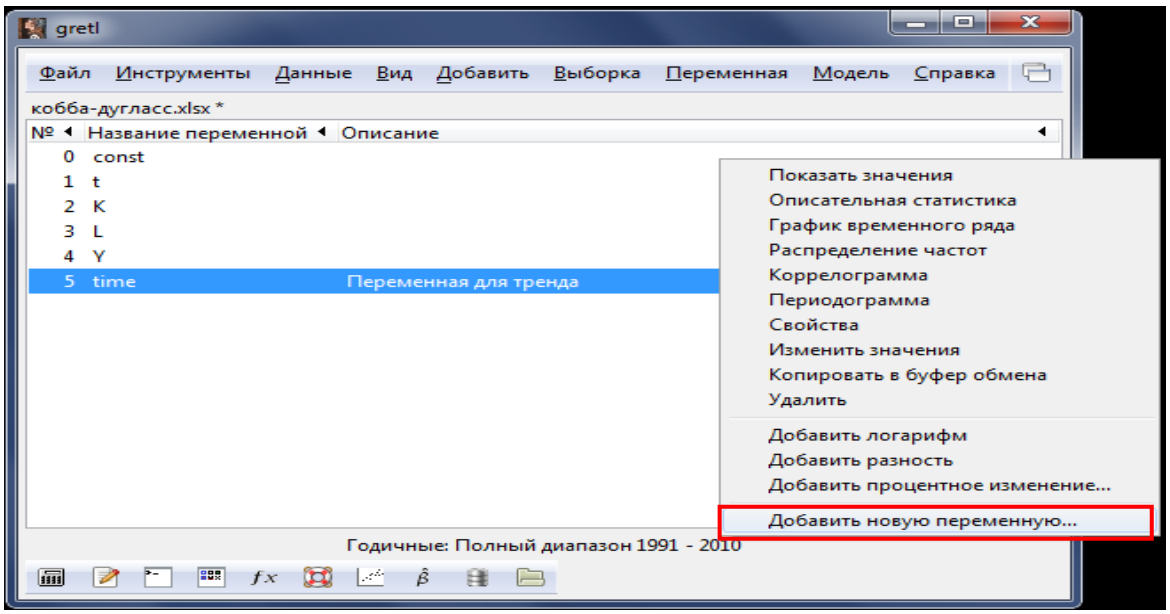


Рисунок 9 – Добавление новой переменной в ППП GRETL

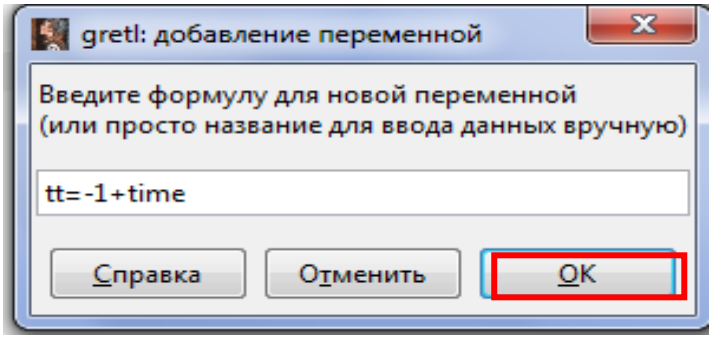


Рисунок 10 – Заполнение диалогового окна *Gretl: добавление переменной*

Добавляем логарифмы переменных *Y*, *K* и *L* (рисунок 11). Затем проделываем цепочку команд *Модель* ⟶ *Метод наименьших квадратов*. Заполняем диалоговое окно согласно рисунку 12⟶*OK*. В результате получаем уравнение линейной регрессии между $\ln Y$ и $\ln K, \ln L, (t - t_0)$ (рисунок 13). Согласно рисунку 13 полученное уравнение таково:

$$\ln Y = 2{,}22 + 0{,}23 \ln K + 0{,}24 \ln L + 0{,}0145(t - t_0).$$

Коэффициент детерминации $R^2 = 99\%$, что говорит о том, что изменение ВВП на 99% обусловлено действием входящих в уравнение факторов – труда и капитала, а 1% изменения ВВП объясняется действием неучтённых факторов.

Уравнение регрессии в целом является статистически значимым, так как расчётный уровень значимости менее 0,05: *P-значение (F)* = $5{,}95 \cdot 10^{-43}$.

Стандартная ошибка уравнения 0,000279. Это говорит о том, что оценивая значение зависимой переменной по найденному уравнению регрессии при фиксированном значении независимых переменных, в среднем будем ошибаться на 0,000279 единиц измерения ВВП. Все коэффициенты регрессии являются статистически значимыми, поскольку расчётные уровни значимости для них не превышают 0,05.

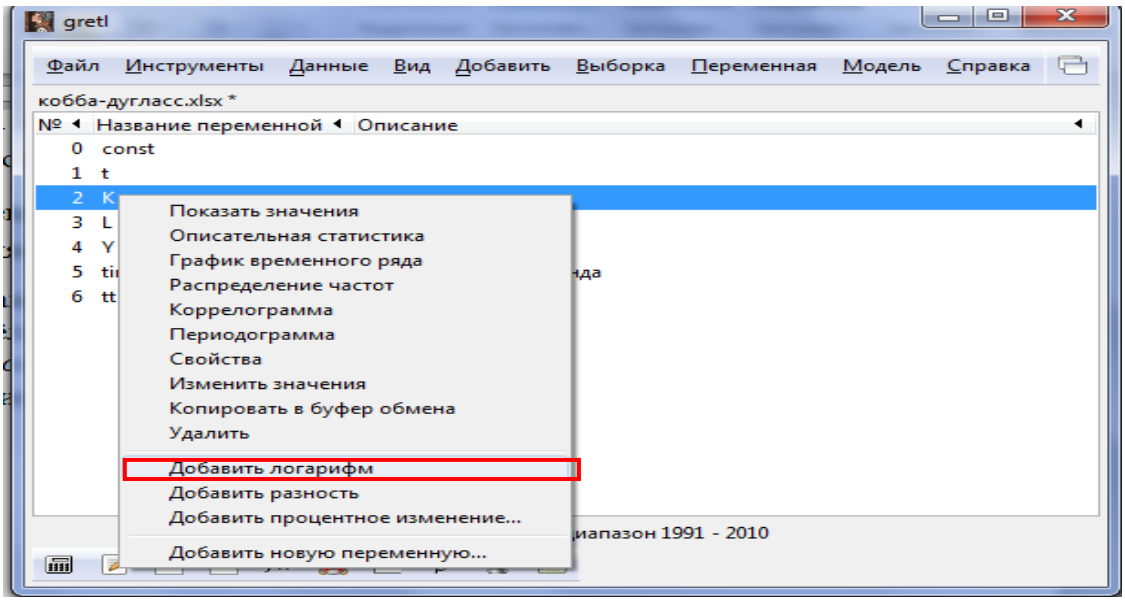


Рисунок 11 – Добавление логарифмов переменных в ППП GRETL

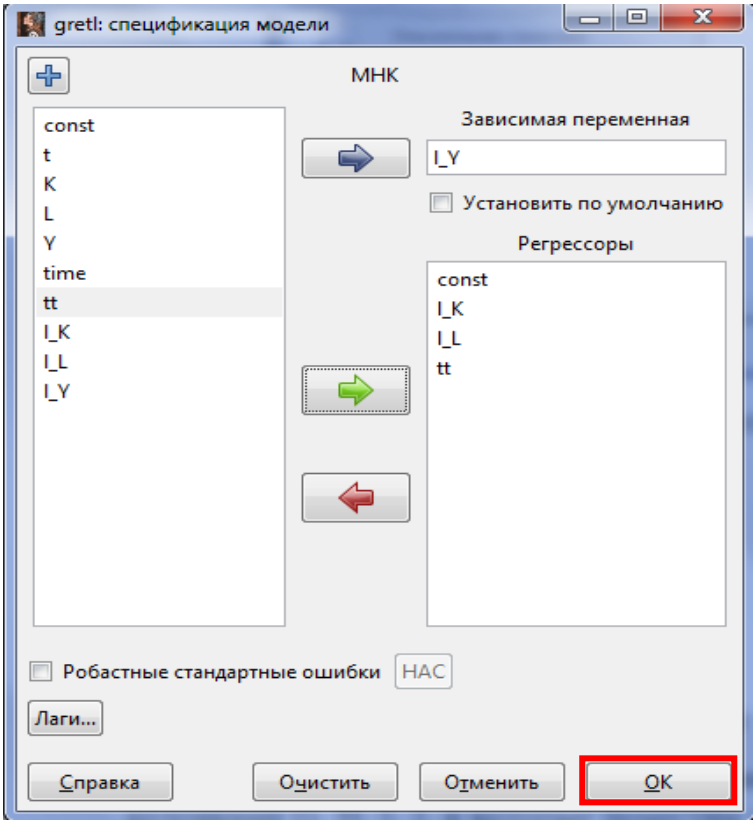


Рисунок 12 – Заполнение диалогового окна метода наименьших квадратов в ППП GRETL

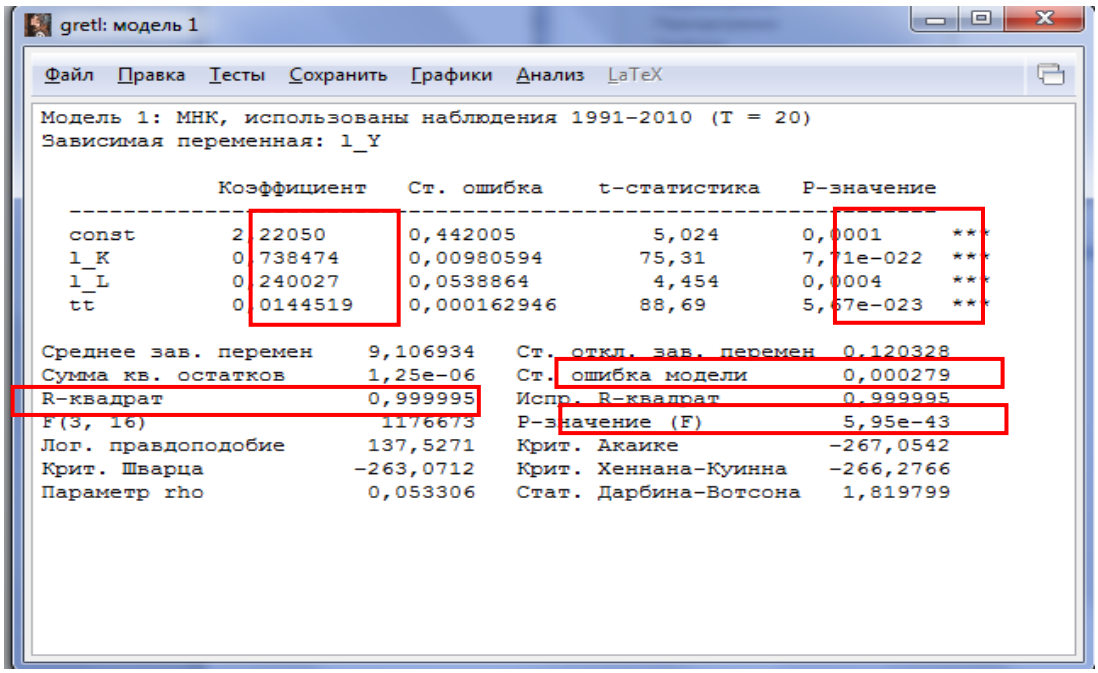


Рисунок 13 – Расчётные значения параметров уравнения регрессии в ППП GRETL.

Из найденного уравнения $\ln Y = 2{,}22 + 0{,}73\ln K + 0{,}24\ln L + 0{,}0145(t - t_0)$ находим параметры $\alpha = 0{,}73$, $\beta = 0{,}24$, $\gamma = 0{,}0145$ и $A = e^{2{,}22} = 9{,}212$.

Модель Солоу такова: $Y = 9{,}212 \cdot K^{0{,}73} \cdot L^{0{,}24} \cdot e^{0.0145(t-t_0)}$.

При отсутствии технического прогресса получим следующее уравнение **производственной функции Кобба-Дугласа**:

$$Y = 9{,}2120 \cdot K^{0{,}73} \cdot L^{0{,}24}.$$

2. Расчет ВВП по модели в условиях наличия и отсутствия технического прогресса.

Рассчитаем ВВП на основе модели производственной функции $Y = 9{,}212 \cdot K^{0{,}73} \cdot L^{0{,}24} \cdot e^{0.0145(t-t_0)}$. Для этого создадим новую переменную Ycolou. Для этого в диалоговом окне *gretl: добавление переменной* вводим формулу *Ycolou=9.212*(K^(0.7385))*(L^(0.24))*(2.718^(0.0145*tt))⟶OK* (рисунок 14).

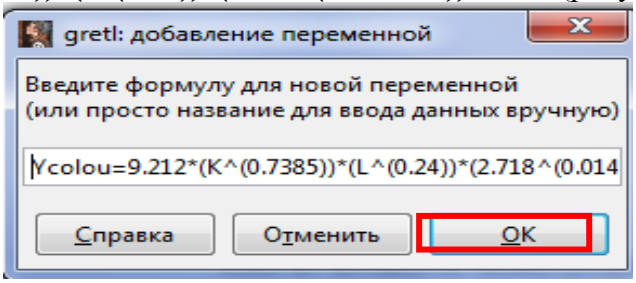


Рисунок 14 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL

Рассчитаем значения ВВП при условии отсутствия технического прогресса, применяя для этого формулу $Y = 9{,}2120 \cdot K^{0{,}73} \cdot L^{0{,}24}$. Создадим новую переменную YKobbaDuglas. Для этого в диалоговом окне *gretl: добавление переменной* вводим формулу YKobbaDuglas=9.212*(K^(0.7385))*(L^(0.24))⟶*OK* (рисунок 15).

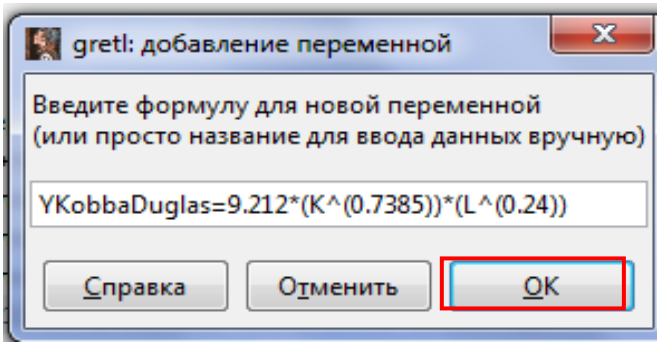


Рисунок 15 – Заполнение диалогового окна *gretl: добавление переменной* в ППП GRETL.

3. Построение графиков.

Построим графические представления временных рядов показателя ВВП (Y), ВВП по модели Солоу (YColou) и ВВП по модели Кобба-Дугласа (YKobbaDuglas). Для этого выбираем *Вид* ⟶ *График* ⟶ *График временного ряда*. Заполняем диалоговое окно согласно рисунку 15 ⟶ *ОК*. Динамики ВВП представлены на рисунке 16.

Из рисунка видно, что значения ВВП, полученные по математической модели Солоу, хорошо согласуются с фактическими значениями ВВП. На графике эти две кривые неразличимы. Существенное различие в значениях и на графике наблюдается при сравнении ВВП с учетом и без учета технического прогресса. Видим, что технический прогресс дает значительное увеличение ВВП.

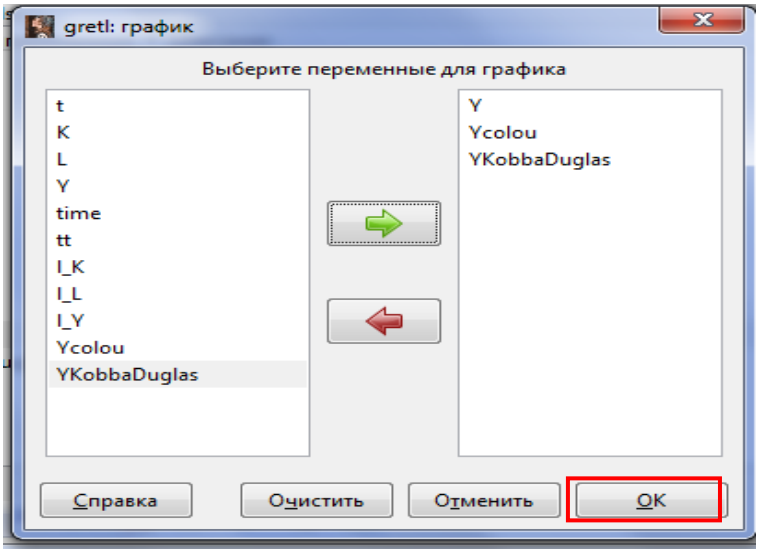


Рисунок 16 – Построение графиков расчётных переменных в ППП GRETL.

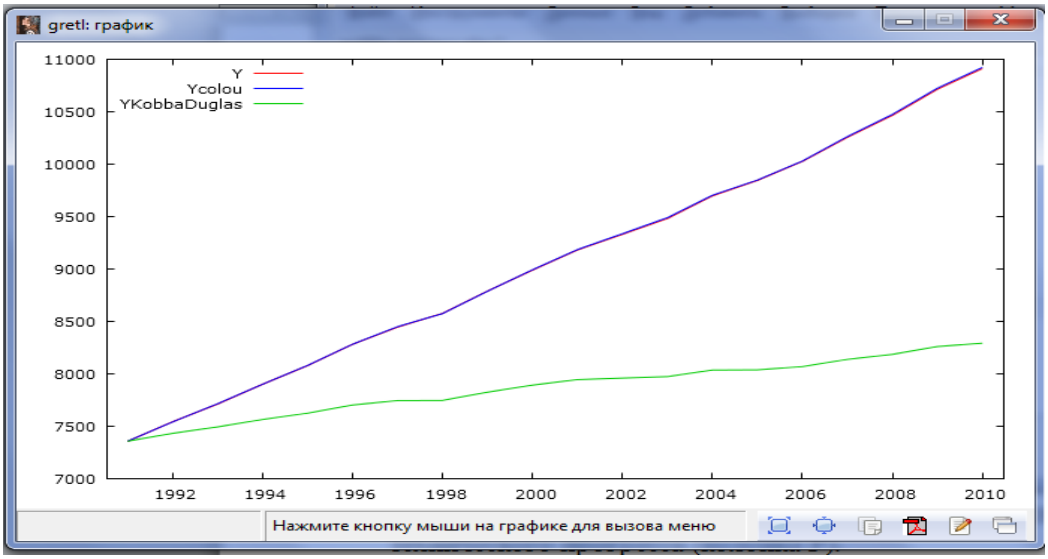


Рисунок 17 – Графики графиков расчётных переменных в ППП GRETL.

4. Нахождение основных характеристик производственной функции. Построим графические представления основных характеристик производственной функции в динамике.

Характеристиками производственной функции Кобба-Дугласа являются:

– средние производительности по капиталу и труду:

$$A_1 = A \cdot K^{\alpha-1} \cdot L^{\beta},\ A_2 = A \cdot K^{\alpha} \cdot L^{\beta-1};$$

– предельные производительности по капиталу и труду:

$$M_1 = A \cdot \alpha \cdot K^{\alpha-1} \cdot L^{\beta},\ M_2 = A \cdot \beta \cdot K^{\alpha} \cdot L^{\beta-1};$$

– частные и общая эластичности:

$$E_1 = \alpha,\ E_2 = \beta,\ E = \alpha + \beta;$$

– технологическая норма замены

$$R_{12} = \frac{\alpha L}{\beta K}.$$

Средняя производительность по каждому ресурсу показывает средний выпуск продукции из расчета единичных затрат данного ресурса.

Предельная или маржинальная производительность по каждому ресурсу показывают приближённо, насколько единиц изменится выпуск, если затраты того или иного ресурса изменятся на единицу:

Частная эластичность по каждому ресурсу приближенно показывает, насколько процентов изменится выпуск, если затраты того или иного ресурса изменятся на один процент. Величина $E = E_1 + E_2$ называется полной эластичностью или ***эластичностью производства***.

Технологическая норма замены называется величина, которая приближенно показывает, как изменится выпуск, если единицу одного ресурса заменить единицей другого.

Для производственной функции Кобба-Дугласа коэффициенты α и β имеют смысл частных эластичностей.

В соответствии с приведенными формулами создадим переменные A1, A2, M1, M2 и R12. Рисунки 18 – демонстрируют создание указанных формул в формате заполнения диалоговых окон.

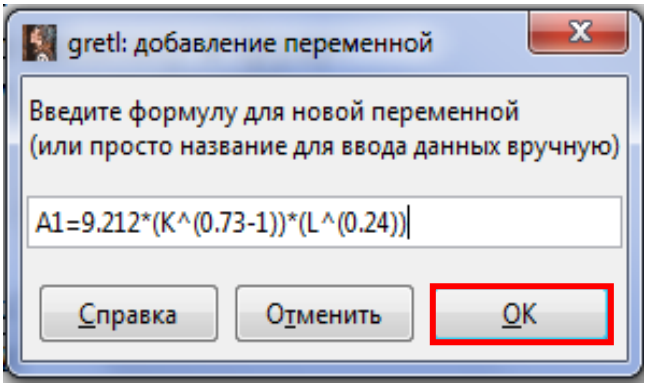


Рисунок 18 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

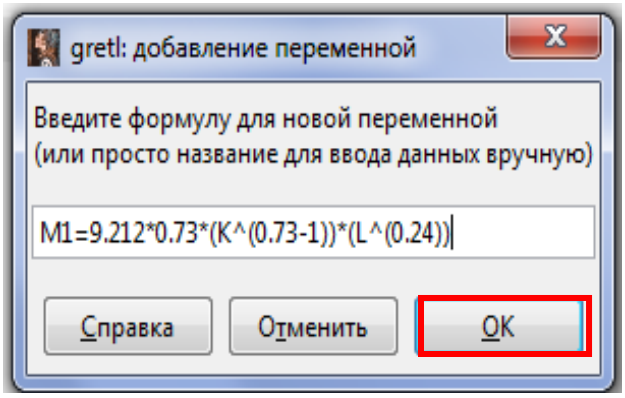


Рисунок 20 – Заполнение диалогового окна *Ggretl: добавление переменной* в ППП GRETL.

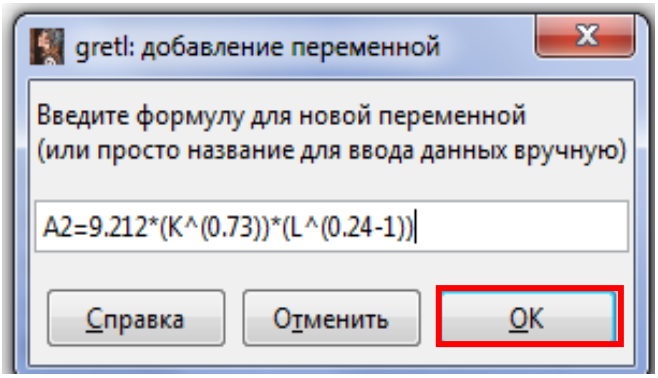


Рисунок 19 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

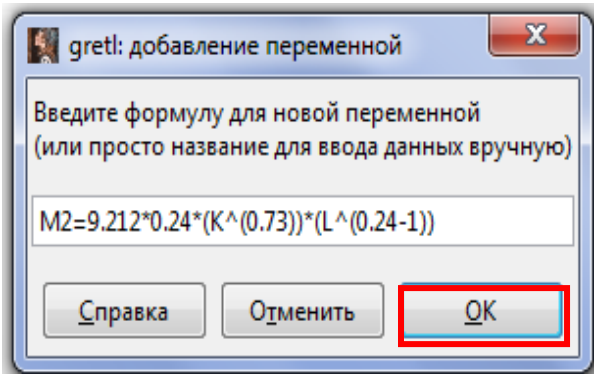


Рисунок 21 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

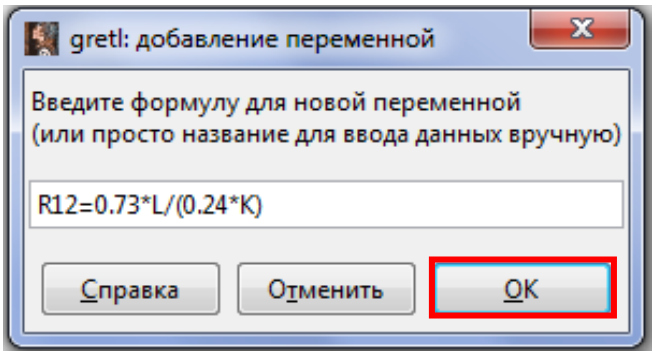


Рисунок 22 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

Динамики средней и предельной производительности по капиталу и труду и технологической нормы замены отражают рисунки 18 – 24. Из рисунков 18 – 23 в

частности. следует, что с течением времени производительности по капиталу убывают, а производительности по труду возрастают.

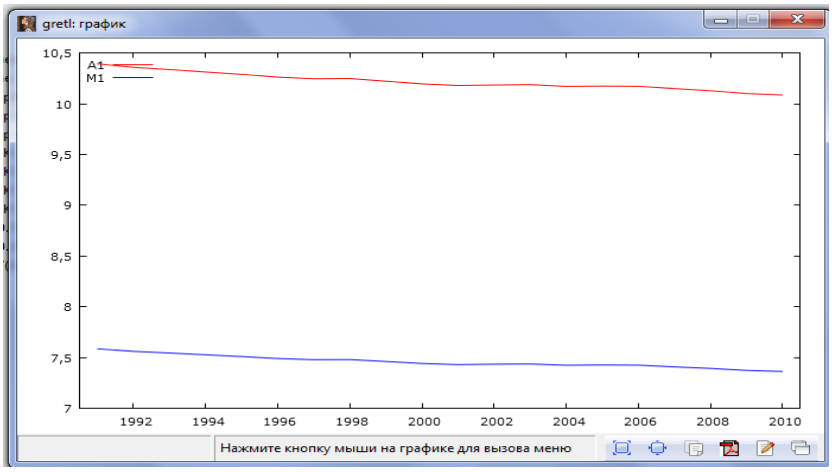


Рисунок 23 – Динамика средней и предельной производительности по капиталу в ППП GRETL.

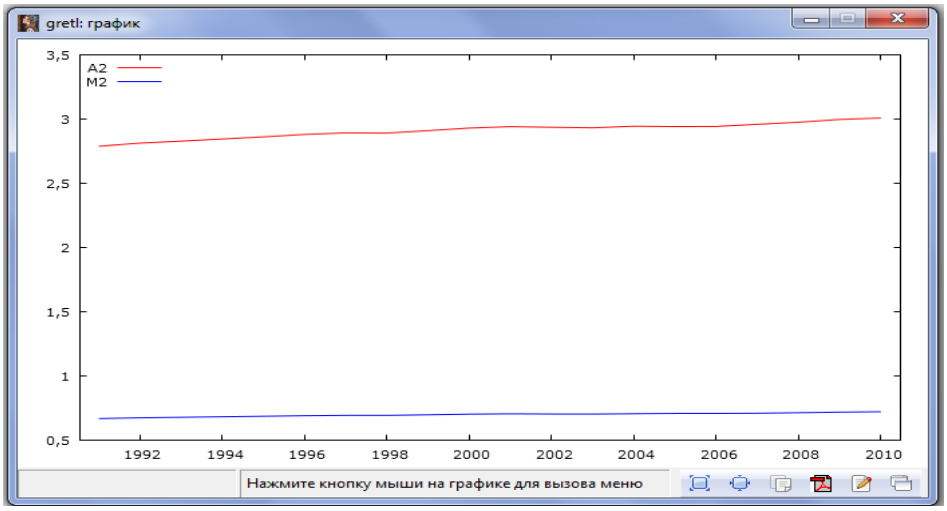


Рисунок 24 – Динамика средней и предельной производительности по труду в ППП GRETL.

Из рисунка видно, что технологическая норма замены данной производственной системы со временем убывает.

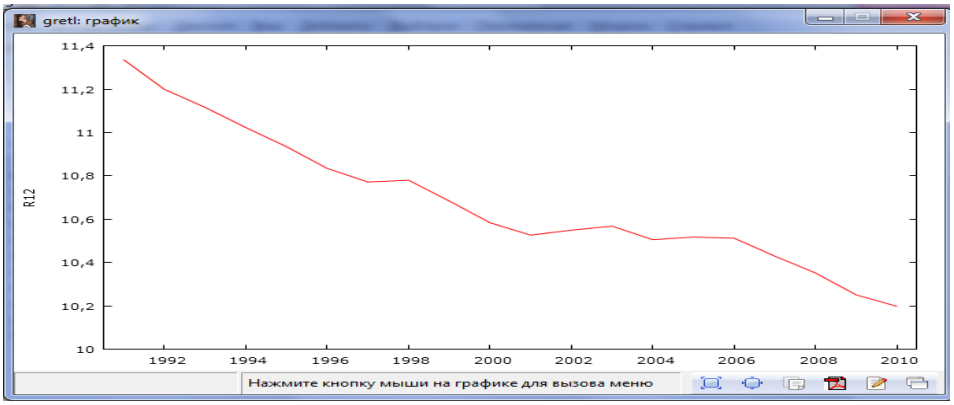


Рисунок 25 – Динамика технологической нормы замены в ППП GRETL.

5. Модель экономической динамики. Построим модель экономической динамики, взяв за основу модель Солоу с линейным и экспоненциальным изменением

нормы накопления.

Наряду с такими показателями, как капитал K_t, труд L_t и выпуск продукции Y_t к основным показателям относятся также фонд накопления S_t и фонд потребления C_t. Эти фонды зависят от нормы накопления s_t за время $t - t_0$.

Обычно рассматривается линейная или экспоненциальная политика изменения нормы накопления, которые имеют вид

$$s_t = s_0 + h \cdot (t - t_0)$$

или $$s_t = s_0 e^{h \cdot (t - t_0)}.$$

Здесь s_0 и h – некоторые постоянные параметры, характеризующие величину нормы накопления. Фонд накопления равен произведению нормы накопления s_t на значение производственной функции $Y_t : S_t = s_t \cdot Y_t$.

Создадим переменные норм накопления st1 и st2, соответствующие линейной и экспоненциальной политике норм накопления соответственно. Рисунки 26 – 27 отражают заполнение диалоговых окон при создании переменных ($s_0 = 0{,}2$, $h = 0{,}02$).

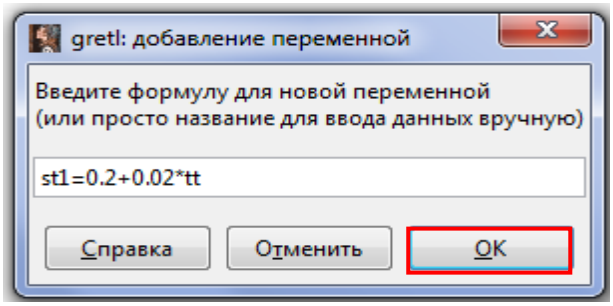


Рисунок 26 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

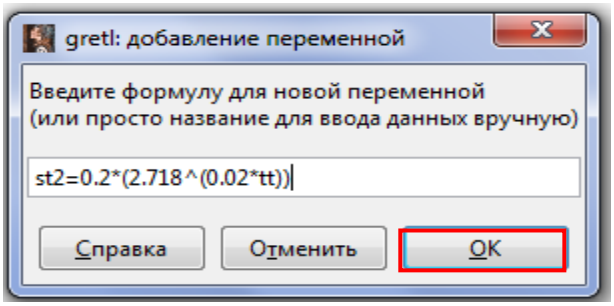


Рисунок 27 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

Фонд потребления равен разности между значением производственной функции и фондом накопления: $C_t = Y_t - S_t$.

К дополнительным показателям относятся:

- фондовооруженность труда K_t / L_t;
- производительность труда Y_t / L_t;
- отдача капитала Y_t / K_t;
- среднедушевое потребление C_t / L_t.

Создадим переменные фондов накопления ST1, ST2 и фондов потребления CT1 и CT2, соответствующие линейной и экспоненциальной политике норм накопления соответственно (рисунки 28 – 31).

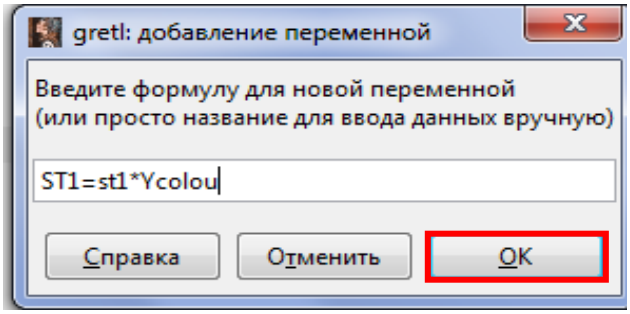


Рисунок 28 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

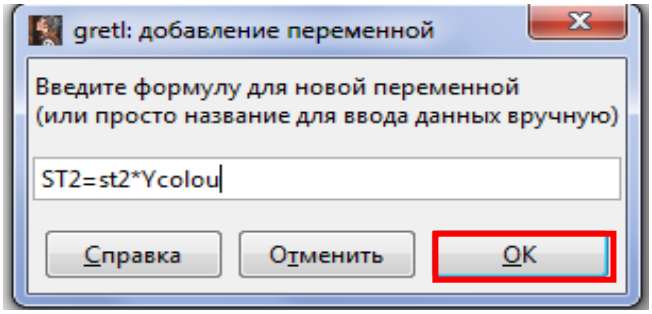


Рисунок 29 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

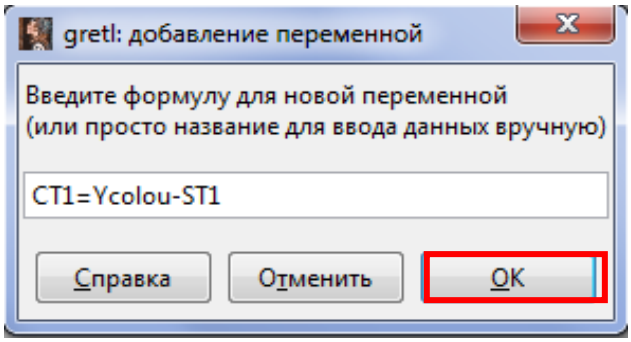


Рисунок 30 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

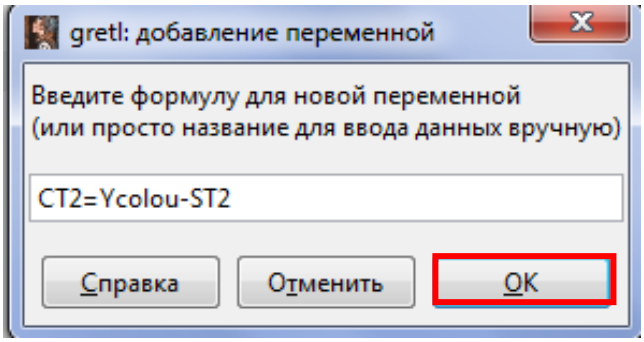


Рисунок 31 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

Изобразим графики изменения фондов накопления и потребления, соответствующие линейной и экспоненциальной политике норм накопления соответственно (рисунки 32 – 33).

Из рисунка 32 видно, что в случае линейной политике норм накопления, фонд накопления растет достаточно быстро, в то время как фонд потребления убывает.

Динамика фондов накопления и потребления для экспоненциальной политики нормы накопления изображена на рисунке 33. Видим существенное изменение соответствующих графиков по сравнению с линейной политикой. В данном случае с течением времени оба фонда возрастают.

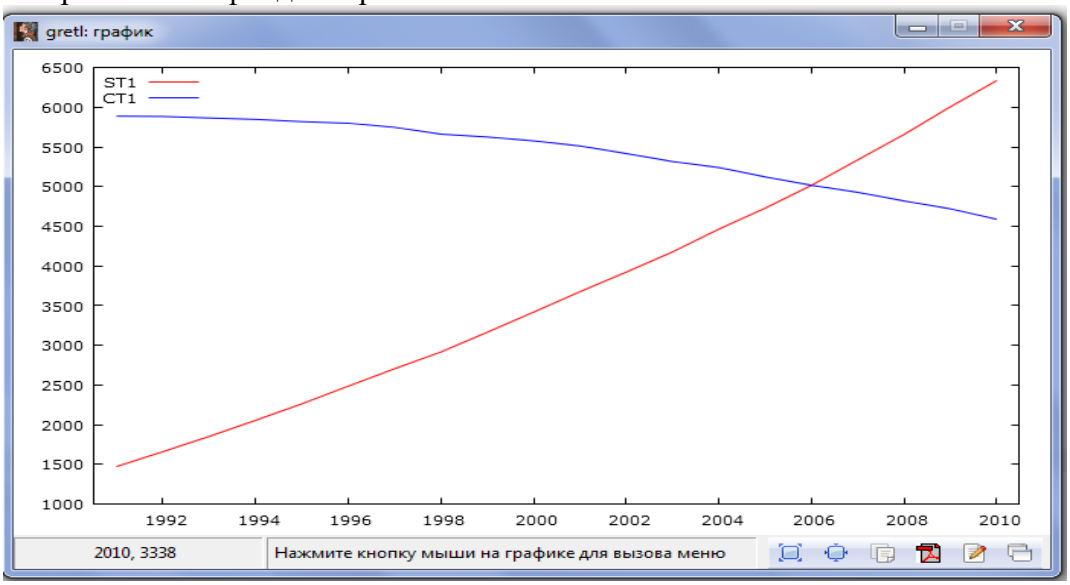


Рисунок 32 – Графики изменения фондов накопления и потребления, соответствующие

линейной политике норм накопления в ППП GRETL.

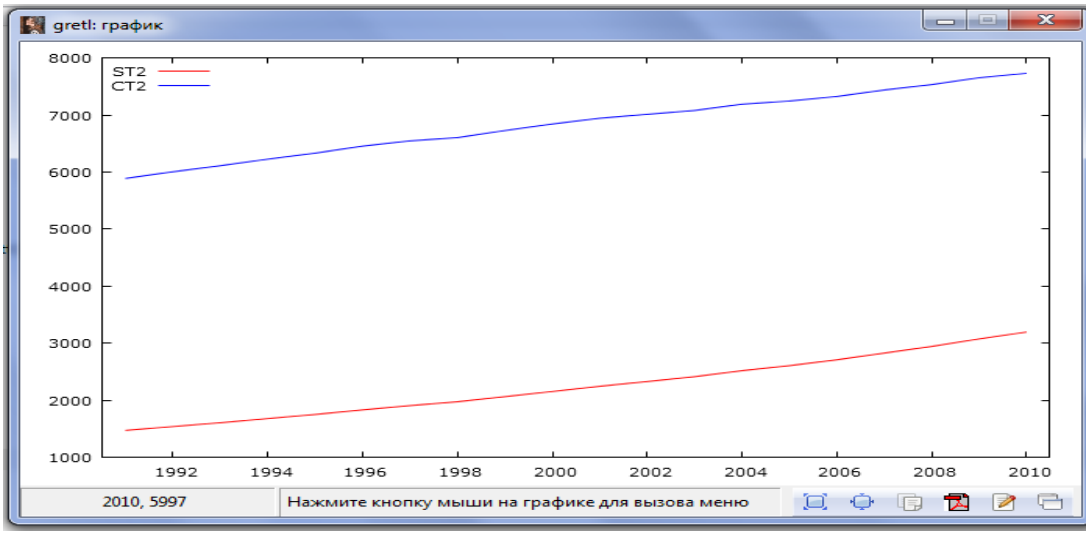


Рисунок 33 – Графики изменения фондов накопления и потребления, соответствующие экспоненциальной политике норм накопления в ППП GRETL.

Создадим переменные дополнительных параметров модели динамики, а именно: фондовооружённость, производительность труда, отдача капитала, среднедушевое потребление. Рисунки 34 – 38 отражают заполнение диалоговых окон при создании переменных.

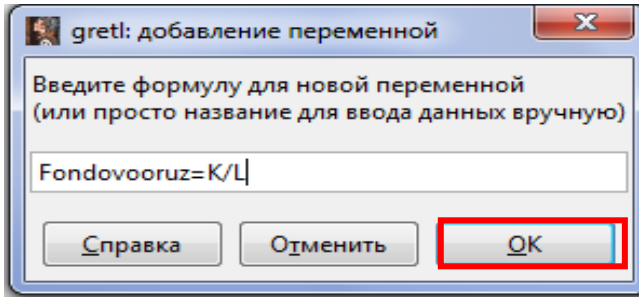


Рисунок 34 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

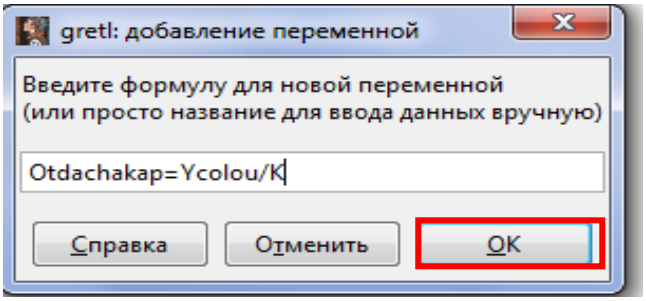


Рисунок 36 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

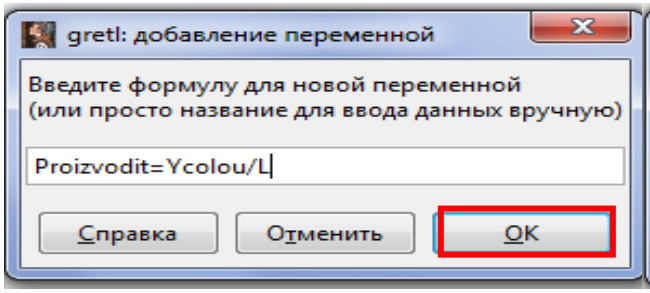


Рисунок 35 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

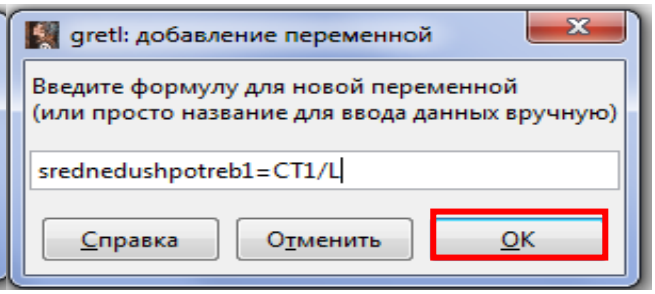


Рисунок 37 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

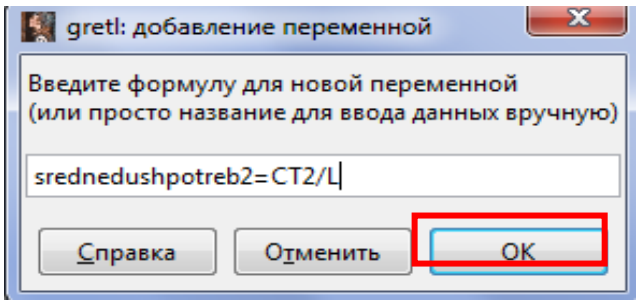


Рисунок 38 – Заполнение диалогового окна *Gretl: добавление переменной* в ППП GRETL.

Изобразим графики динамики дополнительных показателей, полученных в результате проведенных расчетов. На рисунке 39 представлено изменение важнейших дополнительных показателей. Можно сделать вывод, что для данной хозяйственной системы со временем скачкообразно возрастает стоимость основных производственных фондов на одного работника, производительность труда и отдача капитала растут практически линейно.

Критерием успешности развития экономики является показатель среднедушевого (удельного) потребления. На рисунке 40 приведены графики среднедушевого потребления для различных политик нормы накопления. Видим, что среднедушевое потребление выше для экспоненциальной политики нормы потребления. При этом с течением времени этот показатель возрастает.

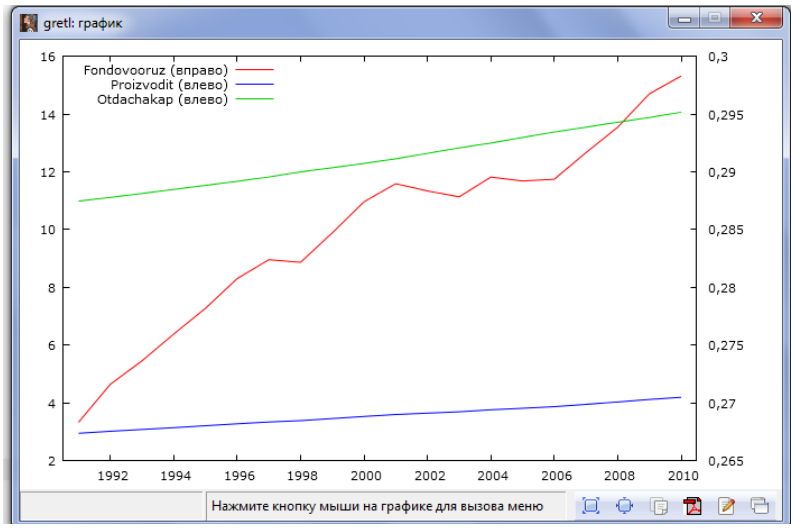


Рисунок 39 – Графики дополнительных параметров модели динамики в ППП GRETL.

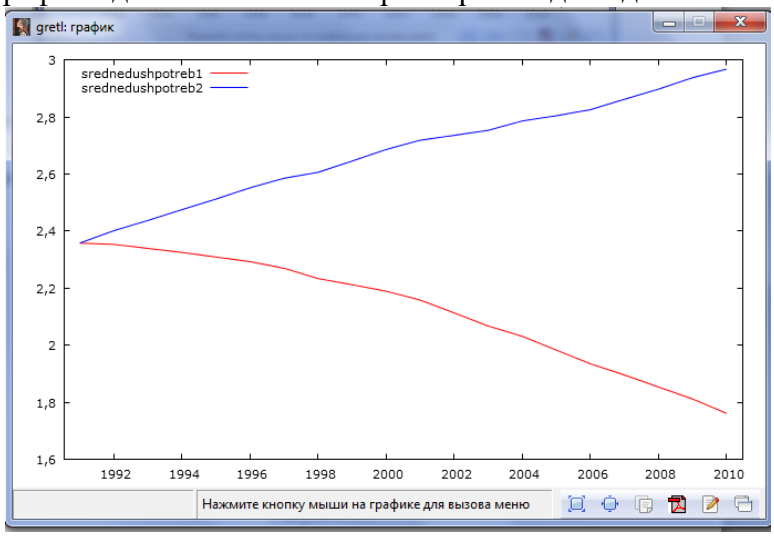


Рисунок 40 – Графики среднедушевого потребления для различных политик нормы накопления в ППП GRETL.

Отчёт по практическому занятию должен содержать:

— постановку задачи (в том числе исходные данные, цель занятия);

— скриншоты расчетных таблиц и графиков, с описанием последовательности выполнения, и полученных промежуточных результатов исследования;

— аналитических оценок, полученных с помощью GRETL с выводами по принятию/ отклонению нулевых гипотез;

— графическое представление средних значений в разрезе уровней факторного признака, экономические выводы из графиков;

— общий вывод по занятию, в том числе итог относительно достижения цели практического занятия.

Типовые оценочные материалы по Теме 2. «Технические и программные средства реализации машинного обучения, возможные проекты применения машинного обучения для оценки рисков»

«Машинное обучение в анализе кредитных и инвестиционных рисков»: Настройка данных для анализа и проектирование модели данных по жилищному кредитованию

Цель занятия: рассмотреть понятие, определение, формы и методы машинного обучения, приобретение навыка регрессионного анализа и оценки корреляции данных средствами машинного обучения.

Методические указания к выполнению практических работ сформулированы в тексте заданий, выдаваемых обучающемуся на занятии.

Ход работы

Задание:

Последовательно выполнить следующие обработки данных.

1. Парциальная предобработка

Парциальная предобработка служит для восстановления пропущенных данных, редактирования аномальных значений и спектральной обработки данных (например, сглаживания данных). Именно эти операции часто проводятся в первую очередь над данными.

Присутствие аномалий при построении моделей оказывает на них большое влияние, ухудшая качество результата.

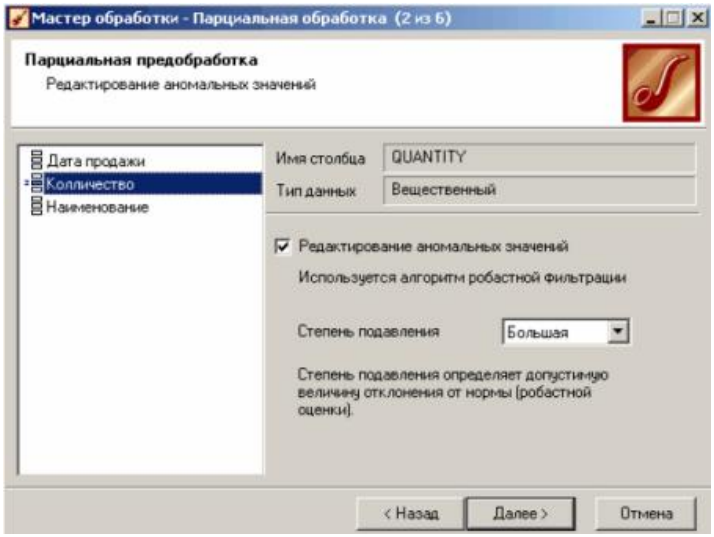
Исходные данные

В качестве примера возьмем данные из файла "Trade.txt". В данном файле находятся данные о продажах за некоторый период.



Обработка данных

Как видно из диаграммы, выбросы ухудшают статистическую картину распределения данных. Воспользуемся Мастером обработки и выберем парциальную обработку.



В Мастере парциальной предобработки на втором шаге выбираем поле "Количество" и указываем ему тип обработки "Редактирование аномальных значений", степень подавления "Большая". Так как больше никаких действий над данными не планировалось, то переходим на шаг запуска процесса обработки и нажимаем "Пуск".

Результат

После выполнения процесса обработки на диаграмме видно, что выбросы уменьшились стала проясняться реальная картина продаж.

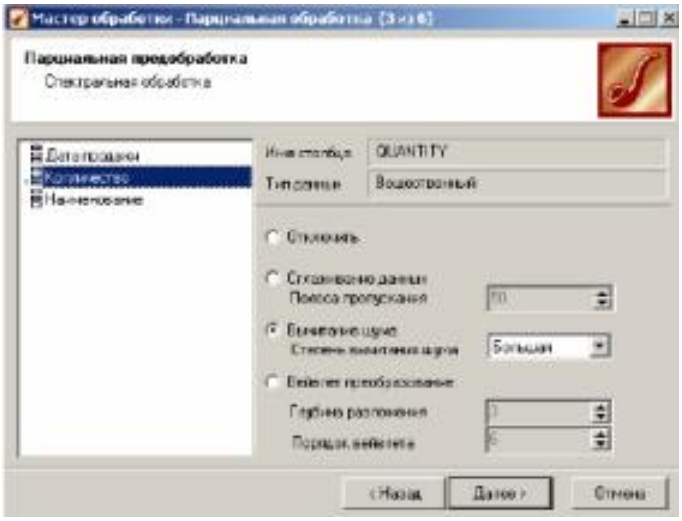


2. Спектральная обработка

Исходные данные

Продемонстрируем такой метод спектральной обработки, как вычитание шума. Для этого продолжим работу с данными файла "Trade.txt". Обработка данных

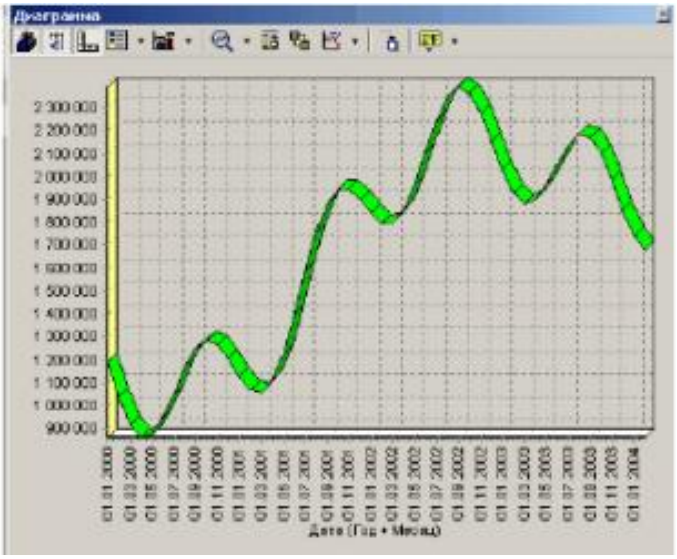
Сгладим данные при помощи парциальной обработки.



В Мастере парциальной предобработки на третьем шаге выбираем поле "Количество" и указываем ему тип обработки "Вычитание шума", степень подавления "Большая". Так как больше никаких обработок не планировалось, то переходим на шаг запуска процесса обработки и нажимаем "Пуск".

Результат

После выполнения процесса обработки выберем в качестве визуализации диаграмму.



Как видно из примера данные стали более сглаженными и могут служить для дальнейшей обработки. Взглянув на данные легко понять общую тенденцию.

3. Корреляционный анализ

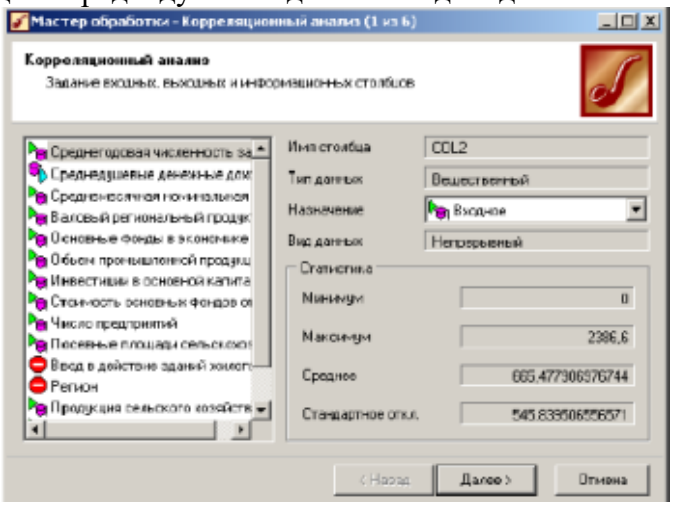
Корреляционный анализ применяется для оценки зависимости выходных полей данных от входных факторов и устранения незначащих факторов. Принцип корреляционного анализа состоит в поиске таких значений, которые в наименьшей степени коррелированны (взаимосвязаны) с выходным результатом. Такие факторы могут быть исключены из результирующего набора данных практически без потери полезной информации. Критерием принятия решения об исключении является порог значимости. Если корреляция (степень взаимозависимости) между входным и выходным факторами меньше порога значимости, то соответствующий фактор отбрасывается как незначащий.

Исходные данные

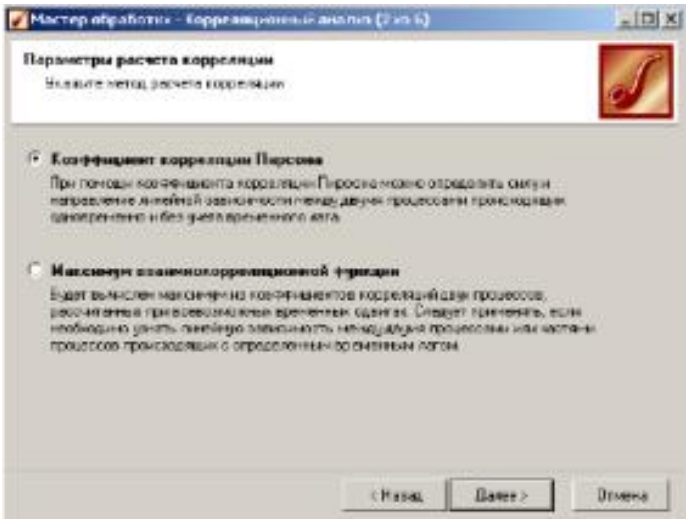
Рассмотрим применение обработки на примере данных из файла "region.txt". В данном примере определим степень влияния экономических показателей региона на среднедушевой доход жителей.

Устранение незначащих входных факторов

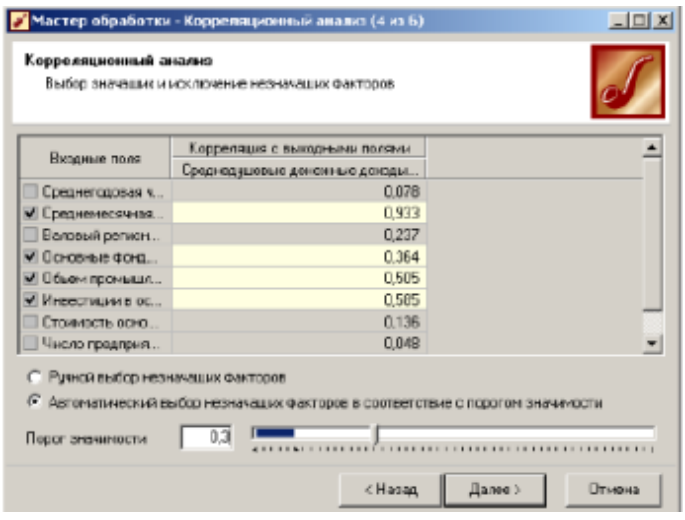
В Мастере обработки выбираем корреляционный анализ и задаем входные и выходные поля. Входными факторами будут являться все экономические показатели региона, а выходным будет "Среднедушевой денежный доход".



На следующем шаге необходимо выбрать метод на основе которого будет происходить расчет *коэффициентов корреляции*, выберем метод коэффициент корреляции Пирсона.



После выполнения предварительных настроек запускаем процесс *корреляционного анализа*, по результатам которого предлагается выбрать, какие факторы оставить для дальнейшей работы. Это делается либо вручную, основываясь на значениях матрицы ковариации, либо путем указания порога значимости (по умолчанию порог значимости равен 0.05).



Результат

По полученной матрице корреляции видно, какие факторы влияют сильнее, чем другие, и какие можно не учитывать при построении всевозможных моделей.

Типовые оценочные материалы по Теме 3. «Основные методы машинного обучения»

«Машинное обучение в анализе кредитных и инвестиционных рисков»: Настройка данных для анализа и проектирование модели данных по жилищному кредитованию

Цель занятия: рассмотреть понятие, определение, формы и методы машинного обучения, используя сведения, представленные на ресурсе Yandex Cloud (обзор сведений о машинном обучении): https://yandex.cloud/ru/blog/posts/2022/10/machine-learning? .

Ход работы

Реализуем первый этап: Настройка данных для анализа и проектирование модели данных

Концепция Big Data – это серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов, которыми сложно управлять с помощью традиционных средств, в виду их структурных различий и большой скорости прироста.

Благодаря технологии Big Data у финансового сектора появилось множество возможностей удобного анализа данных и прогнозирования, например, на основании финансовых сведений по клиенту, предлагать потребителю максимально подходящий продукт, разрабатывать новые продукты, проводить маркетинговые кампании, отслеживать действия конкурентов, в данной работе – оценка объемов жилищного кредитования.

Машинное обучение (Machine Learning) — это способ обучать компьютеры без программирования и явных инструкций, используя только шаблоны и логические выводы. В данной ситуации машине говорят не «сделай 1+1», а показывают пример и просят сделать «похожее». До какой-то степени это схоже с логикой поведения человека. Благодаря развитию ML-подходов машины начинают «делать похожее» с более глубоким уровнем проработки.[1]

Долгое время компьютеры применялись для решения задач, с которыми человек мог справиться и сам. Но постепенно становилось понятно, что преимущество машин — в скорости выполнения операций, быстрой аналитике больших объёмов данных, которые человек вручную обрабатывал бы очень долго. Компьютер может быстро дать результат, нужно только задать правильные условия «на входе». Так появились технологии машинного обучения и искусственный интеллект в широком его понимании.

Часто искусственный интеллект обучается на примерах того, как эта задача решалась человеком (или самой машиной). Наборы таких примеров называются датасетами. Используя датасеты, разработчики обучают алгоритмы классифицировать информацию, строить прогнозы, выделять важные моменты, находить новые пути решения известных задач. Со временем датасеты растут, алгоритмы обучения тоже меняются.

Благодаря развитию технологий машинного обучения Яндекс Музыка научилась самостоятельно находить и добавлять в плейлист песни, которые могут вам понравиться. Сервис анализирует ваши предпочтения, а ещё предпочтения миллионов других пользователей, вкусы которых могут быть похожи на ваши или отличаться. Это развитие технологии, которую, когда-то придумал Netflix и которая полностью изменила логику потребления контента в современном мире.

С такой задачей никогда не справился бы человек — к каждому пользователю невозможно приставить эксперта-музыковеда, который знаком со всем музыкальным

[1] Информация с ресурса Yandex Cloud https://yandex.cloud/ru/blog/posts/2022/10/machine-learning?

наследием человечества и умеет подбирать произведения под разные вкусы. Пример, рекомендательная система Яндекс Музыки - это нетрудно благодаря методам машинного обучения и огромной скорости выполнения операций.

Виды машинного обучения

Для ML-проектов часто требуется высокопроизводительная IT-инфраструктура. Чтобы не тратить время на её создание, настройку и последующее обслуживание, можно использовать облачный сервис Yandex DataSphere. В нём есть необходимые инструменты и динамически масштабируемые ресурсы для полного цикла разработки машинного обучения. Сервис доступен всем пользователям платформы Yandex Cloud. Подробнее о возможностях сервиса рассказали в этом видео:

Перечислим основные виды машинного обучения.

Обучение с учителем

Подход, при котором машине заранее дают понять, какой ответ будет считаться правильным. Под учителем здесь понимается обучающая выборка с размеченными нужным образом данными либо тот, кто указывает правильные ответы в процессе обучения. На основе предоставленных учителем данных машина сама находит и определяет признаки, позволяющие ей, условно, отличить кошку от собаки или мужчину от женщины. В дальнейшем обученный алгоритм можно перенастроить на решение других задач.

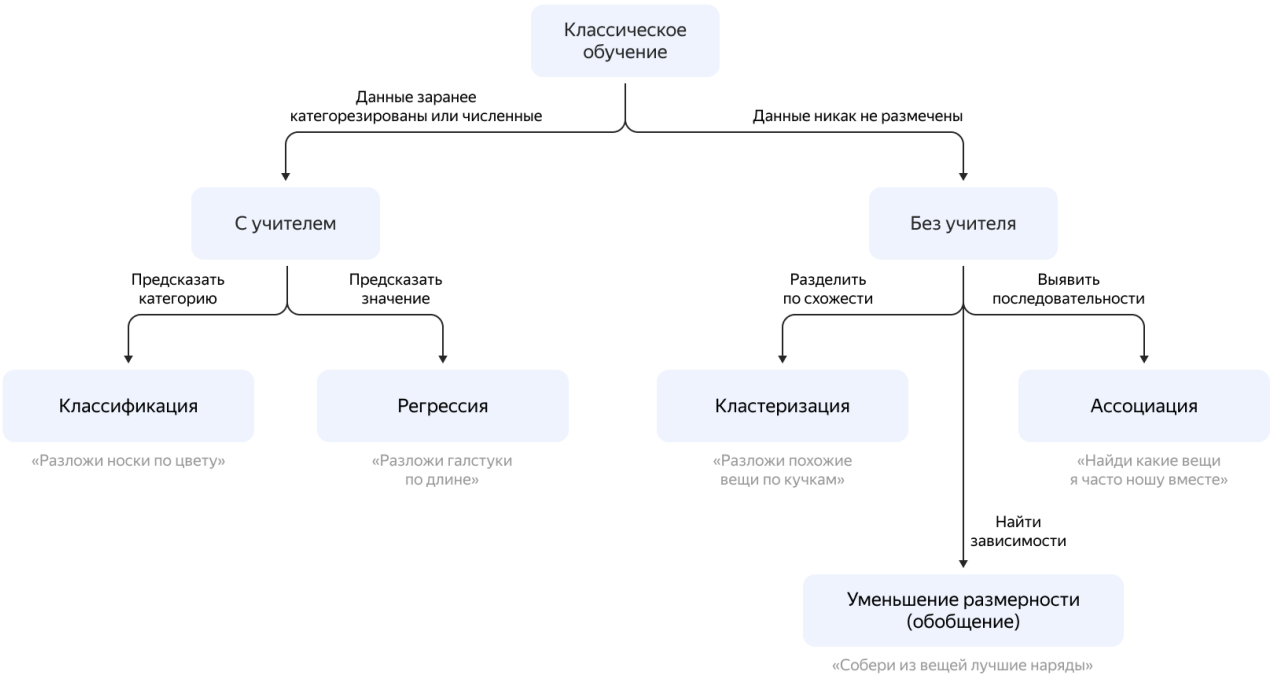


Рис. 41. Виды обучения*

*Источник: ресурс Yandex Cloud https://yandex.cloud/ru/blog/posts/2022/10/machine-learning?

Обучение без учителя

Подход, который актуален ввиду большого количества неразмеченных, хаотично собранных данных. Машина должна сама найти связи между отдельными данными, выявить закономерности, подобрать шаблоны и отсортировать объекты в загруженном датасете. Похожие алгоритмы используются рекомендательными системами интернет-магазинов.[2]

Например, Яндекс Маркет предлагает покупателям товары, которые способны их заинтересовать.

[2] Информация с ресурса Yandex Cloud https://yandex.cloud/ru/blog/posts/2022/10/machine-learning?

Обучение с подкреплением

Более сложный подход, при котором модель обучается сама методом проб и ошибок. Задача алгоритма — минимизировать количество ошибок, поэтому последовательность успешных решений и все неправильные решения фиксируются. Так достигается решение задачи наилучшим образом.

Классический пример успешного применения обученной таким способом машины — победа компьютера IBM Watson в игре Jeopardy!

Возможно, самый сложный уровень обучения искусственного интеллекта на данный момент. Машинное и глубокое обучение часто путают. Их относят к отдельным направлениям развития искусственного интеллекта. Но Deep Learning — это подвид машинного обучения, а нейронные сети — подвид глубокого обучения. Глубокое и машинное обучение различаются способами обучения алгоритмов.

Deep Learning — это способ обозначения более глубоких методов обучения с огромным количеством признаков. Например, задумайтесь, как человек отличает кошку от собаки? У кошки есть усы, другая форма ушей, мордочки и так далее. Каждый из этих параметров — по сути признак. Машина учится также, пытаясь на разных примерах найти признаки, по которым одно можно отличить от другого. Этих признаков может быть огромное количество, и этот параметр отличает глубокое обучение от «обычного».

Можно сказать, что модель машинного обучения — это своеобразный чёрный ящик. ИИ предоставляют условия задачи, и он даёт ответ на выходе. Что в этот момент происходило «под капотом» — неизвестно.

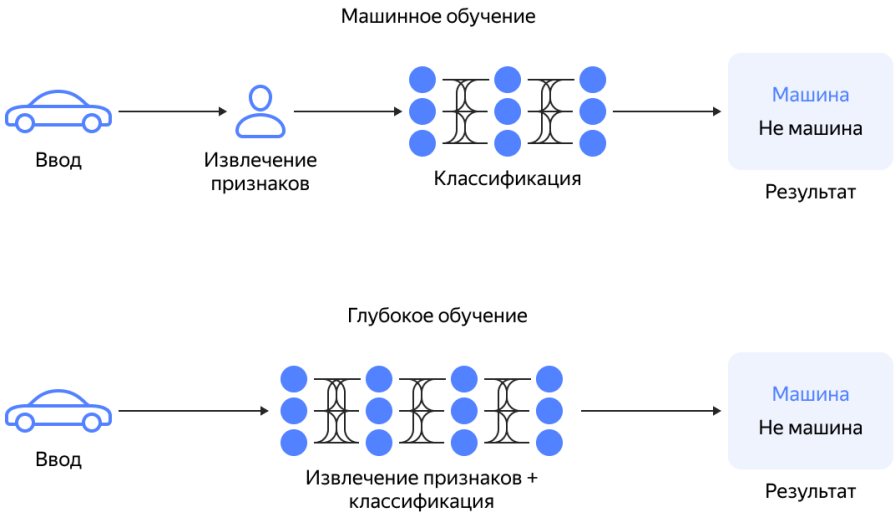


Рис. 42. Виды машинного обучения*

*Источник: ресурс Yandex Cloud https://yandex.cloud/ru/blog/posts/2022/10/machine-learning?

Машинное обучение выглядит как поиск условий, при которых обучаемая модель в подавляющем большинстве случаев начинает выдавать корректные ответы. Для Deep Learning используют огромные датасеты и выделяют много ресурсов. Именно модели нейросетей с глубоким обучением сделали возможным появление систем биометрии, чат-ботов и голосовых ассистентов.

Ещё можно вспомнить и так называемые «ансамбли» — несколько алгоритмов машинного обучения, которые объединены в группу и исправляют ошибки друг друга. Ансамбли применяются в компьютерном зрении, поисковых системах, для технологий распознавания лиц.[3]

Как устроено машинное обучение

Существуют три составляющие любого процесса машинного обучения:

[3] Информация с ресурса Yandex Cloud https://yandex.cloud/ru/blog/posts/2022/10/machine-learning?

- Данные. Это примеры решений, статистика, расчёты и другая информация, которая помогает обучать нейросети. Большие наборы данных называют датасетами.
- Признаки. Это те вещи, на которые модель должна обратить внимание во время обучения. Чем конкретнее признаки и меньше их количество, тем быстрее проходит обучение. У сложных моделей могут использоваться миллионы параметров.
- Алгоритмы. Имеются в виду способы решения задач. У задачи может быть несколько вариантов решения, а машина должна выбрать лучший из них.

Важным свойством машинного обучения является наличие математической связи между входными и выходными данными. Нейросеть не имеет данных об этой взаимосвязи, но может получить их, используя загруженные в неё наборы данных.

Допустим, есть множество однотипных задач, в которых известны условие и правильный ответ (или несколько ответов). Это может быть набор изображений качественных и некачественных деталей на производстве. Нейросеть принимает на вход условие задачи — выбраковку некачественных деталей, пропускает данные сквозь чёрный ящик и даёт на выходе какой-то ответ. Например, данные о количестве плохих деталей.

В модель можно загружать дополнительные параметры, которые будут влиять на обработку входных данных. Машинное обучение строится так, чтобы добиться максимальной частоты корректных ответов на выходе из модели путём подбора параметров. Когда определены нужные параметры, можно заставить машину верно решать аналогичные задачи, даже если она не знает правильных ответов к ним.

Для ML-проектов обычно нужны структурированные наборы данных и высокопроизводительная IT-инфраструктура. Например, созданная Яндексом GPT-подобная нейросеть YaLM на 100 млрд параметров обучалась 65 дней на 1,7 ТБ текстов из интернета, книг и множества других источников с помощью 800 GPU-карт NVIDIA A100. Чтобы не тратить время на создание, настройку и последующее обслуживание большого количества оборудования, можно использовать облачный сервис Yandex DataSphere. В нём есть необходимые инструменты и динамически масштабируемые ресурсы для полного цикла разработки машинного обучения. Сервис доступен всем пользователям платформы Yandex Cloud.

Задание:

1) на основе представленных выше сведений составить список тезаурусов: ключевых понятий «машинного обучения»;
2) разработать теоретическую модель машинного обучения для выбранной финансовой операции:
    а) модель оценки кредитных рисков финансовой организации организации;
    б) модель машинного обучения для прогнозирования на фондовом рынке;
    в) модель подбора инвестиционного портфеля для физического лица;
    г) модель подбора инвестиционного портфеля для юридического лица;
    д) модель анализа суммарных налоговых поступлений по физическим лицам в бюджет в период времени (месяц / год / три года);
    е) модель анализа суммарных налоговых поступлений по юридическим лицам в бюджет в период времени (месяц / год / три года);
    ж) модель анализа расходов регионального бюджета (год / три года);
    з) модель оценки срока окупаемости вновь созданной организации малого бизнеса (в зависимости от видов деятельности);
    и) модель прогнозирования суммарного сбора страховых премий (по видам страхования);
    к) модель оценки рисков страхования (по видам страхования).
3) Далее необходимо проанализировать построенную модель на наличие ошибок и верифицировать ее.

Отчёт по практическому занятию должен содержать:

— постановку задачи (в том числе исходные данные, цель занятия);

— графическое представление средних значений в разрезе уровней факторного признака, экономические выводы из графиков;

— сформированную теоретическую модель машинного обучения, согласно задания практического занятия;

— общий вывод по занятию, в том числе итог относительно достижения цели практического занятия.

Типовые оценочные материалы по Теме 4. «Нейросетевые технологии анализа, прогнозирования финансовых показателей и оценки финансового риска средствами машинного обучения»

Обзор методов машинного обучения, использованных для построения системы бюджетирования. Настройка данных для анализа и проектирование модели данных по бюджетированию, алгоритмы, используемые в машинном обучении

Цель занятия: научиться строить нейросетевую модель анализа данных для системного анализа финансовых показателей на основе аналитической платформ – KNIME, как одного из многих инструментов нейросетевого анализа и программирования.

Задание:

1. Рассмотреть описание методов и форм, использованных для создания фактической модели машинного обучения программистами на базе «1: С»
2. Классификация и описание методов анализа данных.

Термин «машинное обучение» имеет конкретное значение. Методы машинного обучения применяются в тех случаях, когда для решения задачи невозможно подобрать алгоритм, либо, когда алгоритм очень сложен с вычислительной точки зрения.

Например, методы машинного обучения часто применяются в задачах компьютерного зрения и распознавания рукописного текста. Ведь невозможно придумать четкий алгоритм, который будет из массива числовых значений пикселей изображения вычленять области, соответствующие, допустим, лицам людей. Но методы машинного обучения можно применять и в других областях, например, в учете и финансах. Машинное обучение в данных областях решает задачи быстрого расчета и прогнозирования[4].

Модель машинного обучения — это специального вида математическая функция, которая должна аппроксимировать искомую зависимость (мы помним, что вид зависимости нам не известен, либо очень сложен, но мы понимаем, что зависимость есть). У этой функции есть большое количество параметров. Количество параметров может быть от нескольких десятков до миллионов и миллиардов. Для того, чтобы модель могла спрогнозировать данные с достаточной точностью, параметры её необходимо настроить – найти для них оптимальные значения.

В процессе настройки параметров используются исторические данные. Это процесс называется обучением модели. Для обучения моделей придуманы специальные математические алгоритмы. После того как модель обучена, ее можно использовать для расчета и прогнозирования. В общем, никакой магии тут нет. Есть только специальные математические функции и алгоритмы.

С какими проблемами столкнулись программисты при разработке модели машинного обучения для задач бюджетирования и как их решали

Сбор и обработка данных. В учебных примерах данные для обучения модели обычно представлены в более или менее подготовленном виде. А вся предварительная обработка данных заключается в поиске недостоверных значений и поиске зависимостей между различными признаками. В реальности выборку данных для обучения модели пришлось компоновать самим. Дело в том, что в бюджетной модели значения по каждой

[4] Теоретические сведения получены из обзорной статьи Павла Зюзь, ведущего программиста офис NFP компании Первый Бит: https://nfp2b.ru/2023/07/30/magiya-byudzhetirovaniya-kak-my-razrabatyvali-model-mashinnogo-obucheniya-dlya-zadach-byudzhetirovaniya-i-chto-iz-etogo-vyshlo/

статье хранятся в разных строках, также в разных строках хранятся данные по различным значениям аналитики.

Для модели машинного обучения необходимо, чтобы в одной строке были данные сразу по всем статьям и всем аналитикам, от которых зависит выходное значение.

Пример: выбрана бюджетную модель, где данные введены помесячно. Создали модель для расчета показателя прибыли. Известно, что в бюджетной модели прибыль зависит от цены на продукцию (по всей аналитике), и от расходов, а также от расходов в предыдущем месяце. Все эти данные изначально хранятся в базе 1с (используем как источник данных) в различных строках (см. таблицы 3 и 4). Соответственно необходимо скомпоновать различные строки в одну. Для языка программирования Python разработаны библиотеки, сильно упрощающие работу с данными.

Проблема 1: недостаток данных

На первый взгляд, данных в бюджетной модели много, но после компоновки и отбрасывания недостоверных значений остается всего несколько сот строк данных (конечно, все зависит от размера базы данных).

Входные данные исходный формат

Период	Статья бюджета	Аналитика	Значение
01.01.2022	Цена на продукцию	Оборудование класса 1	10000
01.01.2022	Цена на продукцию	Оборудование класса 2	9000
01.01.2022	Цена на продукцию	Запасные части	1500
01.01.2022	Расходы		3000
01.12.2021	Расходы		2800

Таблица 3. Пример входных данных в исходном формате

Входные данные обработанные

Период	Цена на продукцию оборудование 1	Цена на продукцию оборудование 2	Цена на продукцию запчасти	Расходы	Расходы предыдущий период
01.01.2022	10000	9000	1500	3000	2800

Таблица 4. Пример входных данных после обработки

Даже при недостатке данных мы поможем получить приемлемый результат.

Проблема 2: недостоверные, неполные данные или данные разных диапазонов

Выяснилось, что на качество обучения модели влияют ошибки в данных. Каждая ошибка, каждое нерассчитанное значение ухудшает качество обучения модели. По поводу разных диапазонов можно сказать следующее. Например, данные по одному из ЦФО (центр финансовой ответственности) по одной из статей бюджета данные были по ошибке введены в 1000 раз большими, чем нужно. Все подчиненные данные были корректно посчитаны, и мне казалось, что все должно работать, ведь все зависимости сохранены. Но на самом деле модель машинного обучения хорошо работает только в некотором диапазоне данных. А небольшие порции данных, резко выходящие за диапазон остальных данных, очень снижают качество обучения модели.[5]

Следовательно, перед разработкой моделей машинного обучения необходимо очень тщательно проверять данные. Для этого обязательно нужно привлекать специалистов в предметной области (в нашем случае в бюджетировании).

Какие модели выбрать?

Теперь перейдем к вопросу выбора модели. Начали исследование с линейной модели. Предварительное исследование данных показало, что большинство зависимостей

[5] Теоретические сведения получены из обзорной статьи Павла Зюзь, ведущего программиста офис NFP компании Первый Бит: https://nfp2b.ru/2023/07/30/magiya-byudzhetirovaniya-kak-my-razrabatyvali-model-mashinnogo-obucheniya-dlya-zadach-byudzhetirovaniya-i-chto-iz-etogo-vyshlo/

действительно линейные. Пример линейной зависимости *f(x, y) = a*x + b*y +c*. Для таких зависимостей применима линейная модель. Но иногда встречаются зависимости *f(x, y, a) = a*x + b*y +c*, где входными данными кроме *x, y* также является *a*. Такая зависимость уже не является линейной (рисунки 43 и 44). В итоге в качестве модели машинного обучения была выбрана нейронная сеть. Хотя для некоторых случаев я вернулся к линейной модели.

Сравнивая нейронную сеть и квадратичную модель можно сказать, что квадратичная модель в отличие от нейронной сети работает не для всех зависимостей бюджетной модели, но там, где она работает, она может выдать более точный прогноз на большем диапазоне изменений входных значений.

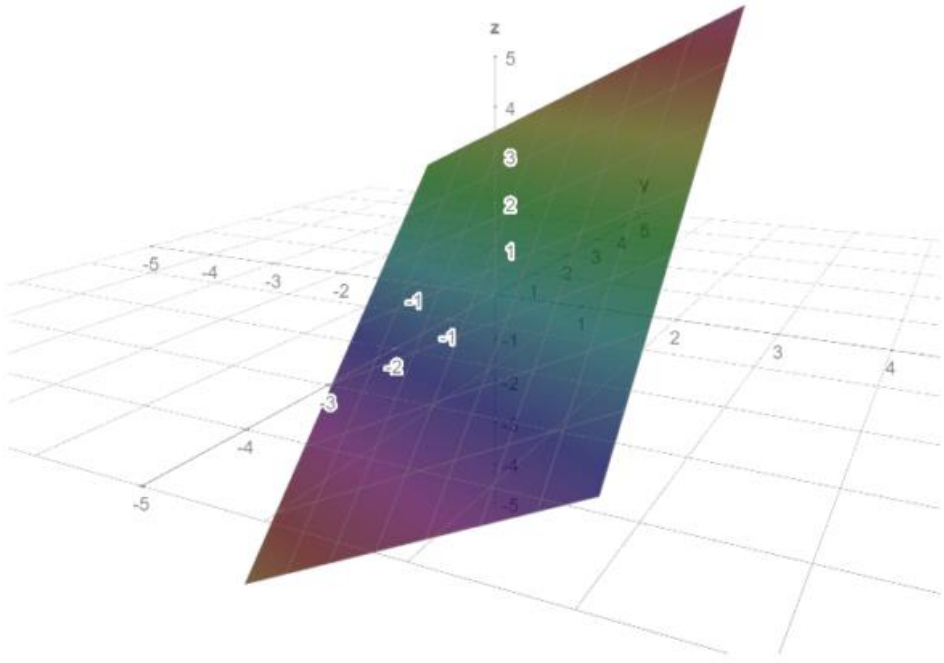


Рисунок 43. Линейная зависимость z = x + 2y

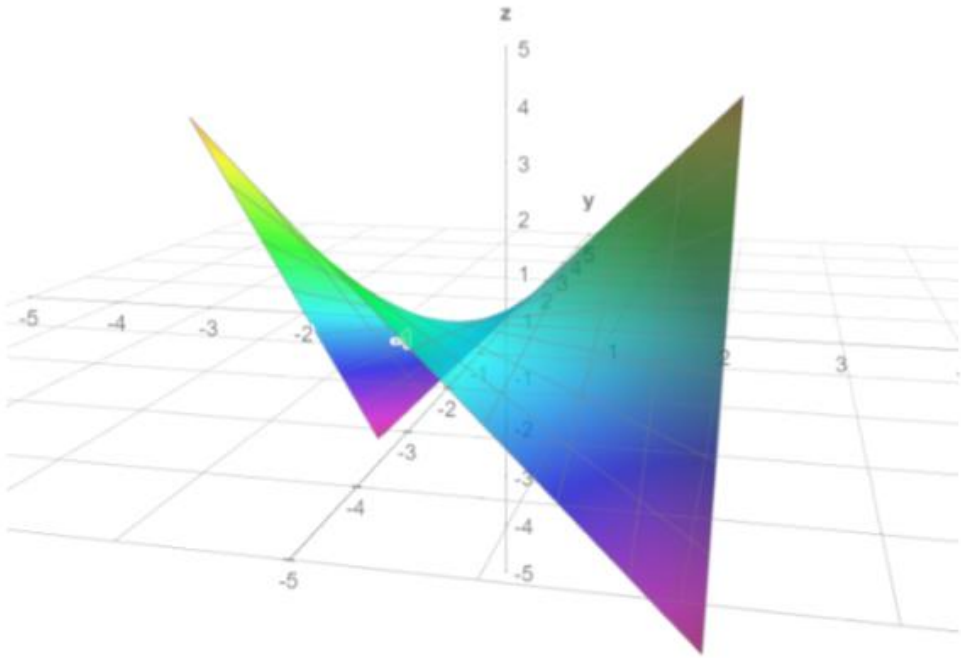


Рисунок 44. Нелинейная зависимость z = xy

Одним из преимуществ методов машинного обучения является скорость. Самой затратной по времени является обучение модели.[6] В рассматриваемом модуле, обучение вынесено в отдельные фоновые процессы. Работа с обученной моделью требует секунды времени (взамен часов и дней при непосредственном расчете по формуле). Большая скорость расчета обеспечивается самой структурой модели.

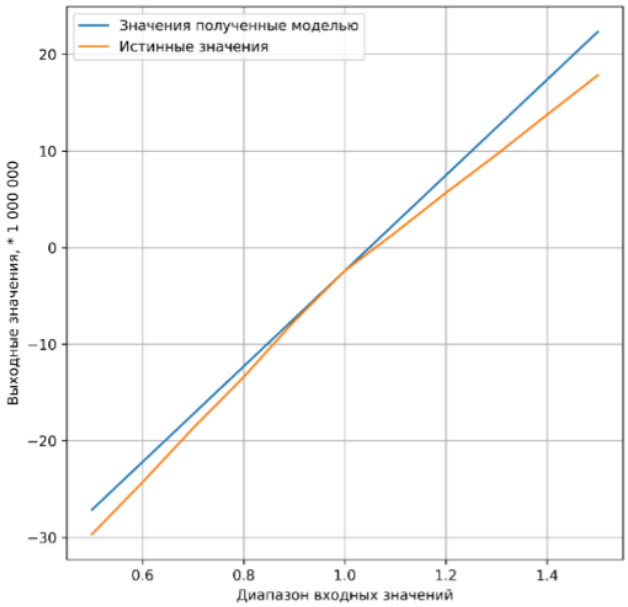
Но, кроме того, несмотря на то, что код моделей был написан на языке python, библиотеки для машинного обучения во многом написаны на C, C++ (известно, что код на python работает медленнее чем на C, C++). Также некоторые библиотеки могут использовать видеокарту для расчетов.

Проблема 3: критерии качества предсказания моделей

Также есть проблема критериев качества. Есть отдельная теория для расчета различных коэффициентов, характеризующих качество обучения модели. Мы выбрали средний процент ошибки в качестве критерия качества. Но также осталась проблема – какое значение процента ошибки считать минимально достаточным? Скорее всего это должно быть требованием заказчика. Интересно то, что качество модели падает, если входные данные сильно отличаются по значениям от тех данных, на которых обучалась модель. Соответственно можно определить диапазон применимости модели (см. рисунки 45 и 46).

Здесь также можно сказать, что качество одной и той же модели машинного обучения может отличаться в зависимости от того, изменение какого входного показателя мы исследуем. Выяснилось, что если для конкретного входного показателя (например, показатель «Количество продаж») в данных, используемых для обучения, много различных значений, то качество модели большом диапазоне изменения этого показателя хорошее (см. рис. 47, 48).

Если по какому-либо показателю мало различных значений в данных обучения (например, показатель «Скидка продаж»), то есть в истории бюджетной модели значение этого показателя редко менялось, то диапазон изменений этого показателя, при котором модель выдает хороший прогноз небольшой. Фактически в этом случае модель выдает хороший прогноз только по тем данных, по которым она обучалась.[7]



[6] Теоретические сведения получены из обзорной статьи Павла Зюзь, ведущего программиста офис NFP компании Первый Бит: https://nfp2b.ru/2023/07/30/magiya-byudzhetirovaniya-kak-my-razrabatyvali-model-mashinnogo-obucheniya-dlya-zadach-byudzhetirovaniya-i-chto-iz-etogo-vyshlo/

[7] Теоретические сведения получены из обзорной статьи Павла Зюзь, ведущего программиста офис NFP компании Первый Бит: https://nfp2b.ru/2023/07/30/magiya-byudzhetirovaniya-kak-my-razrabatyvali-model-mashinnogo-obucheniya-dlya-zadach-byudzhetirovaniya-i-chto-iz-etogo-vyshlo/

Рисунок 45. Сравнение данных, рассчитанных моделью и тестовых данных

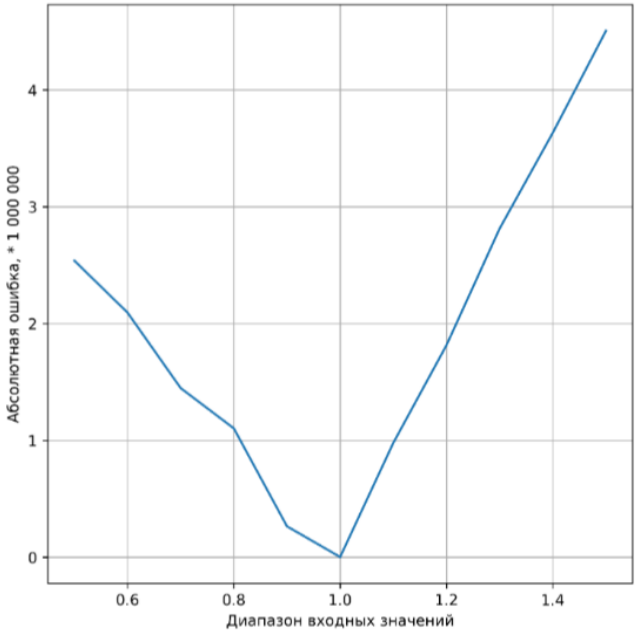


Рисунок 46. Изменение абсолютной ошибки в зависимости от входного значения

Есть еще одна особенность поведения моделей машинного обучения на данных бюджетирования. Есть некоторые показатели, которые мало влияют на выходные данные. Средняя абсолютная и относительная ошибка по таким показателям небольшая, но в этом случае может не сохраняться тренд (по прогнозу при изменении значений такого показателя выходные значения могут увеличиваться, а истинные уменьшаться). Это является допустимым, так как в этом случае истинный диапазон изменения выходного показателя и диапазон изменения прогнозного значения меньше погрешности модели (то есть почти 0).[8]

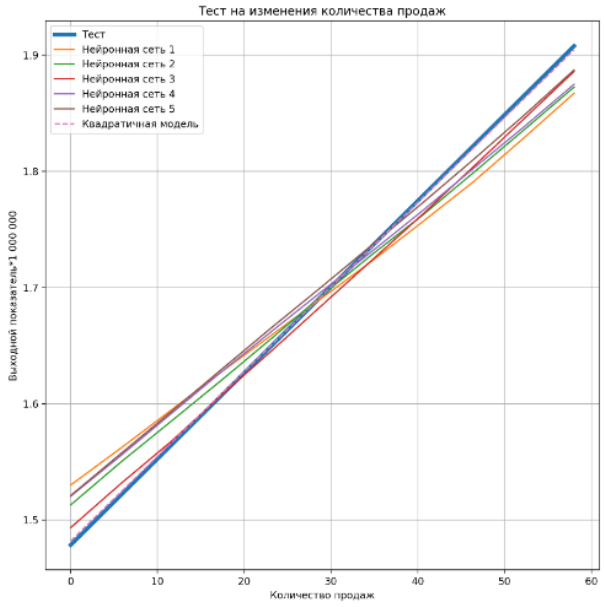


Рисунок 47. Тест прогноза различных моделей при изменении количества продаж

[8] Теоретические сведения получены из обзорной статьи Павла Зюзь, ведущего программиста офис NFP компании Первый Бит: https://nfp2b.ru/2023/07/30/magiya-byudzhetirovaniya-kak-my-razrabatyvali-model-mashinnogo-obucheniya-dlya-zadach-byudzhetirovaniya-i-chto-iz-etogo-vyshlo/

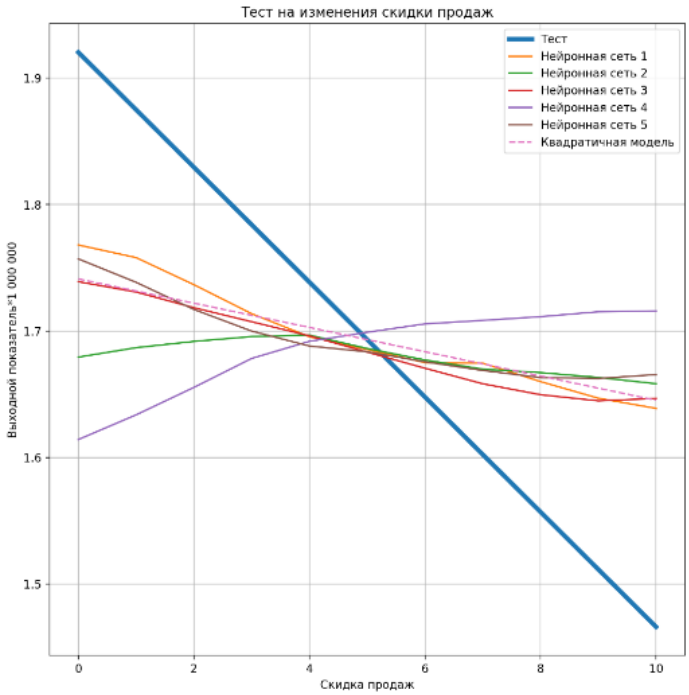


Рисунок 48. Тест прогноза различных моделей при изменении скидки продаж

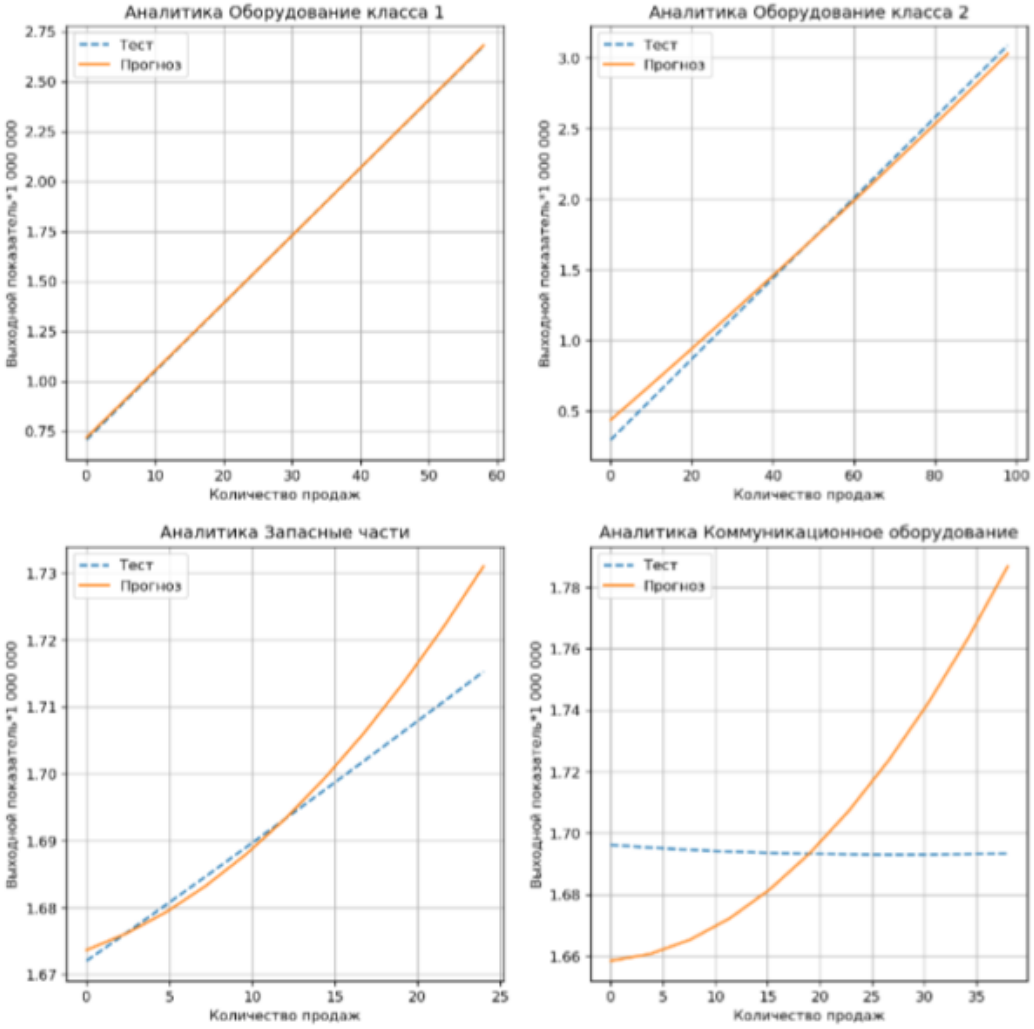


Рисунок 49. Тест прогноза модели по различным аналитикам. По аналитикам «Запасные части», «Коммуникационное оборудование» изменение выходного показателя близко к нулю

Преимущества и недостатки методов машинного обучения.

Плюсы:

- скорость вычисления в режиме эксплуатации
- модели машинного обучения находят скрытые, неочевидные и трудно формализуемые зависимости. Соответственно есть возможность посчитать то, что не удается посчитать иными методами.

Минусы:

- трудности с выбором конкретных моделей машинного обучения

Самое главное — это отсутствие порядка в данных.

Машинное обучение применяется в различных областях, но бизнес пока готовится к применению ML. Тратьте минимум времени и получайте наилучшие результаты для принятия важных решений с помощью VBM.[9]

Отчёт по практическому занятию должен содержать:

— постановку задачи (в том числе исходные данные, цель занятия);

— графическое представление средних значений в разрезе уровней факторного признака, экономические выводы из графиков;

— сформированную теоретическую модель машинного обучения, согласно задания практического занятия;

— общий вывод по занятию, в том числе итог относительно достижения цели практического занятия.

Типовые оценочные материалы по Теме 5 «Средства интернет - технологий структурирования и организации данных»

Цель занятия – закрепление практических навыков работы с ППП MS EXCEL для обработки BigData: использование инструментов дисперсионного анализа пакетов прикладных программ, подключение графических опций с последующим анализом таблиц одно- и двухфакторного дисперсионного анализа, а также графиков средних для главных эффектов и пересечений. Спроектировать модель машинного обучения на основе полученного дисперсионного анализа (в части кластеризации)

Ключевые слова и выражения: вариация, дисперсия, дисперсионный анализ, фактор, однофакторный анализ, двухфакторный анализ, графики средних.

Задание:

Целями работы являются:

- ознакомление со структур ой источников открытых данных, изучение способов хранения и представления данных;

- приобретение навыка построения системы сбора данных.

Задачи:

1 Исследование наборов данных, представленных на портале открытых данных data.gov.ru.

2 Исследование наборов данных, представленных на портале data.worldbank.org.

3 Построение автоматизированной системы сбора данных.

В качестве инструментального средства используется программное обеспечение Microsoft Excel.

Задания по части 1 «Исследование портала data.gov.ru »

1 Государство http://data.gov.ru/rubriki/gosudarstvo

2 Экономика http://data.gov.ru/rubriki/ekonomika

3 Образование http://data.gov.ru/rubriki/education

4 Здоровье http://data.gov.ru/rubriki/zdorove

[9] Теоретические сведения получены из обзорной статьи Павла Зюзь, ведущего программиста офис NFP компании Первый Бит: https://nfp2b.ru/2023/07/30/magiya-byudzhetirovaniya-kak-my-razrabatyvali-model-mashinnogo-obucheniya-dlya-zadach-byudzhetirovaniya-i-chto-iz-etogo-vyshlo/

5 Экология http://data.gov.ru/rubrics/ecology

6 Транспорт http://data.gov.ru/rubriki/transport

7 Культура http://data.gov.ru/rubrics/culture

8 Спорт http://data.gov.ru/rubrics/sport

9 Строительство http://data.gov.ru/rubriki/stroitelstvo

10 Досуг и отдых http://data.gov.ru/rubrics/leisure -and - entertainment

11 Торговля http://data.gov.ru/rubriki/torgovlya

12 Туризм http://data.gov.ru/rubrics/tourism

13 Электроника http://data.gov.ru/rubrics/electronics

14 Картография http://data.gov.ru/rubrics/cartography

15 Безопасность http://data.gov.ru/rubriki/bezopasnost

16 Метеоданные http://data.gov.ru/rubrics/weather

Порядок выполнения

1 Исследование наборов данных на портал е data.gov.ru:

1.1 Выберите вариант задания.

1.2 Найдите произвольный набор данных на портале data.gov.ru по тематике , указанной в выбранном варианте задания. Набор должен быть представлен в формате csv и кодировке Windows.

1.3 Загрузите на компьютер найденный набор данных и его паспорт.

1.4 Проведите анализ набора данных : определите количество записей и полей в наборе данных.

2 Исследование наборов данных на портал е data.worldbank.org:

2.1 Выберите вариант задания.

2.2 Найдите произвольный набор данных на портале data.worldbank.org по тематике , указанной в выбранном варианте задания.

2.3 Загрузите на компьютер найденный набор данных в формате XLS .

2.4 На основе набора данных подготовьте выборку, содержащую значения показателя за все годы для трёх произвольно выбранных стран мира.

2.5 На основе подготовленной выборки постройте график, иллюстрирующий изменение показателя со временем для трёх стран мира.

2.6 Сохраните файл.

3 Построение системы автоматизированной системы сбора данных:

3.1 Выберите вариант задания.

3.2 Найдите интернет-сайт, содержащий указанные в задании данные.

3.3 Запустите Microsoft Excel.

3.4 Выберите пункт «Из Веба» в меню «Данные».

3.5 В адресной строке появившегося окна «Создание веб-запроса» наберите адрес найденной ранее веб-страницы.

3.6 Выберите таблицу, содержащую искомые данные.

3.7 Нажмите кнопку «Импорт».

3.8 В появившемся окне «Импорт данных» нажмите кнопку «Свойства».

3.9 В появившемся окне «Свойства внешнего диапазона» задайте параметр «Период обновления», равны й 1 минуте, параметр «Обновление при открытии файла» - «Да».

3.10 Нажмите кнопку «ОК».

3.11 В окне «Импорт данных» нажмите кнопку «ОК».

3.1 2 Сохраните файл.

4 Отчет о работе:

4.1 Составьте отчет о работе.

4.2 Преобразуйте отчет в формат PDF.

4.3 Запакуйте отчет (PDF) и все использованные и созданные в работе файлы в архив формата ZIP.

Содержание отчета

Отчет должен содержать:

1 Титульный лист: наименование работы, вариант задания, ФИО студента, номер учебной групп ы , дата выполнения работы.

2 Реферат.

3 Оглавление.

4 Часть 1 «Исследование наборов данных на портале data.gov.ru»:

4.1 Задание.

4.2 Копия экрана с набором данных, открытом в Microsoft Excel.

4.3 Описание набора данных согласно нижеприведенной форме

6 Часть 3 «Построение автоматизированной системы сбора данных»:

6.1 Задание.

6.2 Копи я экрана с интернет-сайтом, содержащим данные.

6.3 Копия экрана Microsoft Excel после импорта данных.

7 Список использованных источников:

7.1 Источники данных.

7.2 Нормативные документы.

Все представленные в отчете таблицы и рисунки должны иметь пояснения. Отчет должен быть оформлен в соответствии с действующими стандартами института.

Теме 6. «Системы машинного обучения и нейросетевого анализа в отдельных секторах экономики и производства. Методы машинного обучения при поиске информации в современных системах с помощью систем машинного обучения»

Цель занятия – закрепление практических навыков работы с ППП MS EXCEL для обработки BigData: использование инструментов дисперсионного анализа пакетов прикладных программ, подключение графических опций с последующим анализом таблиц одно- и двухфакторного дисперсионного анализа, а также графиков средних для главных эффектов и пересечений. Спроектировать модель машинного обучения на основе полученного дисперсионного анализа (в части кластеризации)

Ключевые слова и выражения: вариация, дисперсия, дисперсионный анализ, фактор, однофакторный анализ, двухфакторный анализ, графики средних.

Задание:

1. Согласно порядковому номеру в журнале старосты выбрать вариант задачи.
2. Сформулировать нулевые гипотезы для проведения однофакторного и двухфакторного дисперсионного анализа при уровне значимости α=5%.
3. Проверить нулевые гипотезы с помощью ППП MS EXCEL. Получить графическое представление средних значений в разрезе уровней факторного признака.
4. Сформулировать ключевые выводы.
5. Спроектировать модель нейросетевого анализа на основе полученного дисперсионного анализа (в части кластеризации).

Задача о влиянии стажа работы и пола работника на среднюю производительность труда. При исследовании влияния стажа работы на производительность труда (количество деталей в день) в одном из цехов завода получена следующая статистика. Выяснить оказывает ли влияние каждый фактор по отдельности на среднюю производительность труда работника, а также совместное сочетание факторов.

Производительность труда	Стаж работы (лет)	Пол работника	Производительность труда	Стаж работы (лет)	Пол работника
155	До 5	м	154	5-10	м
153	До 5	м	158	5-10	ж
149	До 5	м	157	5-10	м
150	До 5	ж	161	5-10	ж
147	До 5	ж	149	5-10	м

149	До 5	м	150	5-10	м
153	До 5	м	152	5-10	ж
155	До 5	ж	148	5-10	ж
158	До 5	м	159	5-10	м
156	До 5	ж	161	5-10	м
154	До 5	ж	164	5-10	ж
172	До 5	м	158	5-10	ж
170	До 5	ж	163	5-10	ж
169	До 5	м	175	5-10	м
178	До 5	ж	171	5-10	м
169	До 5	ж	174	5-10	ж
153	10-15	м	164	15-20	м
162	10-15	ж	162	15-20	м
164	10-15	ж	163	15-20	ж
163	10-15	м	160	15-20	ж
167	10-15	ж	163	15-20	м
155	10-15	ж	166	15-20	ж
149	10-15	м	168	15-20	м
156	10-15	м	167	15-20	ж
161	10-15	ж	164	15-20	ж
160	10-15	м	166	15-20	м
158	10-15	м	183	15-20	ж
166	10-15	ж	176	15-20	м
163	10-15	ж	182	15-20	ж
165	10-15	м	179	15-20	ж
177	10-15	м	181	15-20	м
183	10-15	ж	180	15-20	м

Для проведения одно- и двухфакторного дисперсионного анализа необходимо сформулировать **нулевые гипотезы**:

H_0^A : влияние стажа работы на среднюю производительность труда отсутствует;

H_0^B : влияние пола работника на среднюю производительность труда отсутствует;

H_0^{AB} : совместное влияние стажа работы и пола работника на среднюю производительность труда отсутствует.

Реализуем инструменты *Однофакторный дисперсионный анализ* и *Двухфакторный дисперсионный анализ с повторениями* в ППП MS EXCEL.

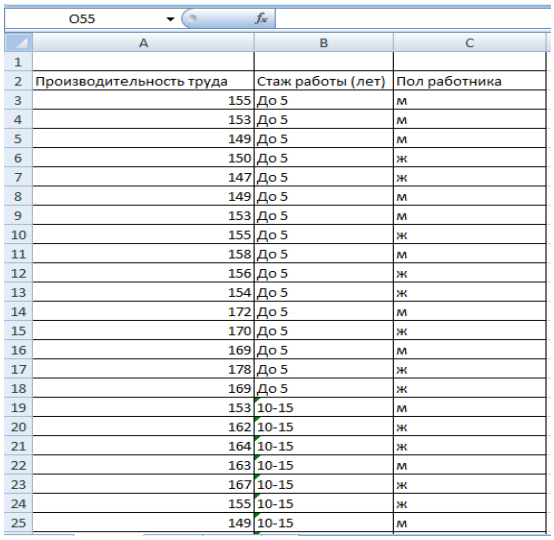
Факторный признак №1 – стаж работы (лет). Уровни факторного признака:

— до 5 лет;
— 5 – 10 лет;
— 10 – 15 лет;
— 15 – 20 лет.

Факторный признак №2 – пол работника. Уровни факторного признака:

— мужской (м);
— женский (ж).

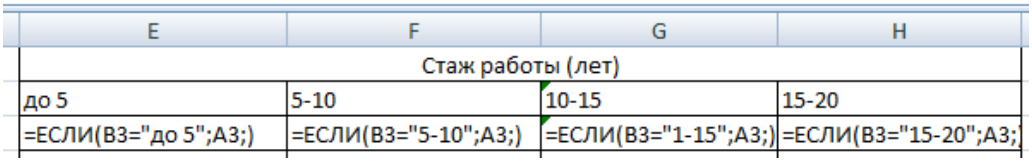
1. Заносим данные на Лист1 MS EXCEL в ячейки *A2:C66* согласно рисунку 1.



	A	B	C
1			
2	Производительность труда	Стаж работы (лет)	Пол работника
3	155	До 5	м
4	153	До 5	м
5	149	До 5	м
6	150	До 5	ж
7	147	До 5	ж
8	149	До 5	м
9	153	До 5	м
10	155	До 5	ж
11	158	До 5	м
12	156	До 5	ж
13	154	До 5	ж
14	172	До 5	м
15	170	До 5	ж
16	169	До 5	м
17	178	До 5	ж
18	169	До 5	ж
19	153	10-15	м
20	162	10-15	ж
21	164	10-15	ж
22	163	10-15	м
23	167	10-15	ж
24	155	10-15	ж
25	149	10-15	м

Рисунок 1 – Исходный массив данных для проведения дисперсионного анализа

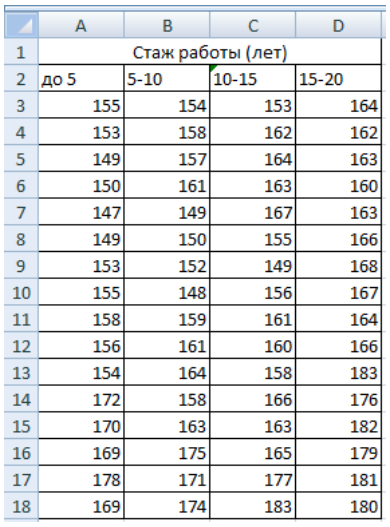
2. Для проведения однофакторного дисперсионного анализа по фактору *Стаж работы* проведём сортировку массива по уровням факторного признака. Для этого в ячейках *E1:H3* введём формулы согласно рисунку 2. Затем скопируем формулы во все ячейки массива *E3:H66*.



E	F	G	H
Стаж работы (лет)			
до 5	5-10	10-15	15-20
=ЕСЛИ(B3="до 5";A3;)	=ЕСЛИ(B3="5-10";A3;)	=ЕСЛИ(B3="1-15";A3;)	=ЕСЛИ(B3="15-20";A3;)

Рисунок 2 – Отсев производительности труда по уровням факторного признака – стаж работы.

Используя настройки фильтра, на Листе2 MS EXCEL можно сформировать таблицу для проведения однофакторного дисперсионного анализа (рисунок 3).



	A	B	C	D
1	Стаж работы (лет)			
2	до 5	5-10	10-15	15-20
3	155	154	153	164
4	153	158	162	162
5	149	157	164	163
6	150	161	163	160
7	147	149	167	163
8	149	150	155	166
9	153	152	149	168
10	155	148	156	167
11	158	159	161	164
12	156	161	160	166
13	154	164	158	183
14	172	158	166	176
15	170	163	163	182
16	169	175	165	179
17	178	171	177	181
18	169	174	183	180

Рисунок 3 – Таблица для проведения однофакторного дисперсионного анализа

2. На панели меню выбираем *Данные* → *Анализ данных*. В диалоговом окне *Анализ данных* выбираем *Однофакторный дисперсионный анализ* (рисунок 4) → *ОК*.

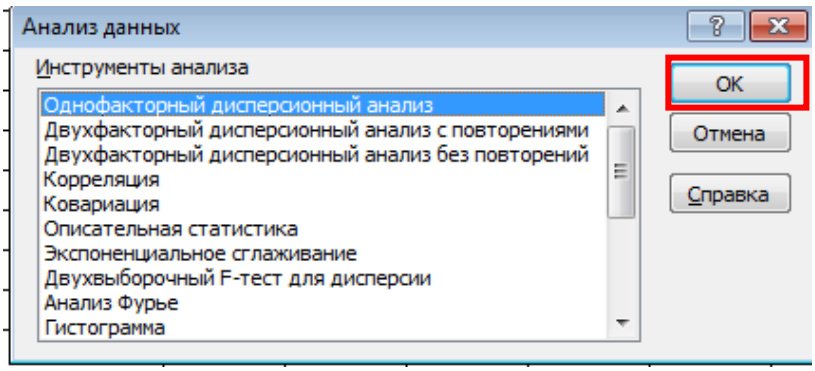


Рисунок 4 – Выбор инструмента анализа *Однофакторный дисперсионный анализ* в ППП MS EXCEL.

3. Диалоговое окно *Однофакторный дисперсионный анализ* заполняем согласно рисунку 5 → *ОК*.

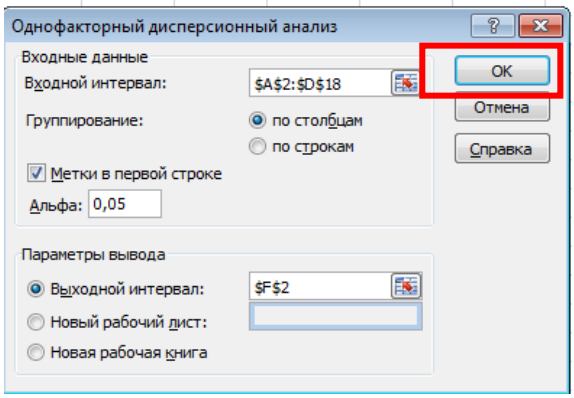


Рисунок 5 – Диалоговое окно *Однофакторный дисперсионный анализ* в ППП MS EXCEL

4. Результаты однофакторного дисперсионного анализа отражает рисунок 6. Поскольку *Р-Значение* равно 0,0014, что существенно меньше заданного уровня значимости $\alpha = 0{,}05$, то нулевую гипотезу H_0^A о независимости влияния стажа работы на среднюю производительность труда отклоняем. Таким образом, **стаж работы оказывает влияние на среднюю производительность труда.**

Однофакторный дисперсионный анализ

ИТОГИ

Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
до 5	16	2537	158,5625	94,1291667
5-10	16	2554	159,625	69,9833333
10-15	16	2602	162,625	71,45
15-20	16	2724	170,25	68,6

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	1337,047	3	445,6823	5,86110769	0,001405231	2,758078316
Внутри групп	4562,438	60	76,04063			
Итого	5899,484	63				

Рисунок 6 – Результаты Однофакторного дисперсионного анализа в ППП MS EXCEL.

5. Для проведения однофакторного дисперсионного анализа по фактору *Пол работника* проведём сортировку массива по уровням факторного признака. Для этого в ячейках *J1:K3* введём формулы согласно рисунку 7. Затем скопируем формулы во все

ячейки таблицы *J3:K66*.

J	K
Пол работника	
м	ж
=ЕСЛИ(C3="м";A3;)	=ЕСЛИ(C3="ж";A3;)

Рисунок 7 – Отсев производительности труда по уровням факторного признака – пол работника.

Используя фильтр для последнего массива, на Лист3 MS EXCEL сформируем таблицу для проведения однофакторного дисперсионного анализа (рисунок 8).

	A	B
1	Пол работника	
2	м	ж
3	155	150
4	153	147
5	149	155
6	149	156
7	153	154
8	158	170
9	172	178
10	169	169
11	153	162
12	163	164
13	149	167
14	156	155
15	160	161
16	158	166
17	165	163
18	177	183
19	154	158
20	157	161
21	149	152
22	150	148
23	159	164
24	161	158
25	175	163

Лист1 Лист2

Рисунок 8 – Таблица для проведения однофакторного дисперсионного анализа

6. На панели меню выбираем *Данные* → *Анализ данных*. В диалоговом окне *Анализ данных* выбираем *Однофакторный дисперсионный анализ* → *ОК*.

Диалоговое окно *Однофакторный дисперсионный анализ* заполняем согласно рисунку 9 → *ОК*.

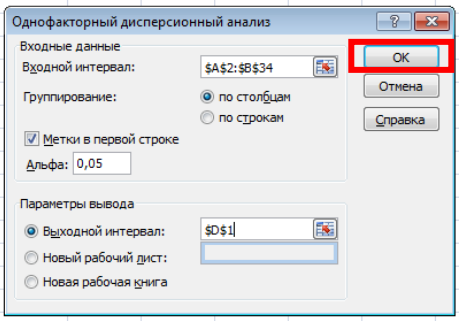


Рисунок 9 – Диалоговое окно *Однофакторный дисперсионный анализ* в ППП MS EXCEL

7. Результаты однофакторного дисперсионного анализа отражает рисунок 10. Поскольку *P-Значение* равно 0,39, что больше заданного уровня значимости $\alpha = 0{,}05$, то

нет оснований отвергнуть нулевую гипотезу H_0^B . Таким образом, **пол сотрудника не оказывает влияние на среднюю производительность труда.**

ИТОГИ

Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
м	32	5175	161,7188	92,01512097
ж	32	5242	163,8125	96,02822581

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	70,14063	1	70,14063	0,746004857	0,391073222	3,995887049
Внутри групп	5829,344	62	94,02167			
Итого	5899,484	63				

Рисунок 10 – Результаты Однофакторного дисперсионного анализа в ППП MS EXCEL

8. Для проведения двухфакторного дисперсионного анализа скопируем исходный массив данных на Лист4 MS EXCEL в ячейки *A2:C66*.

	E	F	G	H	I	J	K	L
1								
2	до 5		5-10		10-15		15-20	
3	м	ж	м	ж	м	ж	м	ж
4	=ЕСЛИ(И(B3="До 5";C3="м");A3;)	=ЕСЛИ(И(B3="До 5";C3="ж");A3;)	=ЕСЛИ(И(B3="5-10";C3="м");A3;)	=ЕСЛИ(И(B3="5-10";C3="ж");A3;)	=ЕСЛИ(И(B3="10-15";C3="м");A3;)	=ЕСЛИ(И(B3="10-15";C3="ж");A3;)	=ЕСЛИ(И(B3="15-20";C3="м");A3;)	=ЕСЛИ(И(B3="15-20";C3="ж");A3;)

Рисунок 11 – Отсев производительности труда в разрезе двух факторов.

Для отсева производительности труда в разрезе двух факторов в ячейки *E2:L4* заносим формулы по аналогии с рисунком 11. Затем скопируем формулы во все ячейки массива *E3:L66*.

1. Используя фильтр для последнего массива, на Лист5 MS EXCEL сформируем таблицу для проведения двухфакторного дисперсионного анализа (рисунок 12).

	A	B	C	D	E
1		**до 5**	**5-10**	**10-15**	**15-20**
2	**м**	155	154	153	164
3		153	157	163	162
4		149	149	149	163
5		149	150	156	168
6		153	159	160	166
7		158	161	158	176
8		172	175	165	181
9		169	171	177	180
10	**ж**	150	158	162	163
11		147	161	164	160
12		155	152	167	166
13		156	148	155	167
14		154	164	161	164
15		170	158	166	183
16		178	163	163	182
17		169	174	183	179

Рисунок 12 – Таблица для проведения двухфакторного дисперсионного анализа

10. На панели меню выбираем *Данные → Анализ данных*. В диалоговом окне *Анализ данных* выбираем *Двухфакторный дисперсионный анализ с повторениями → ОК* (рисунок 13). Диалоговое окно *Двухфакторный дисперсионный анализ с повторениями* заполняем согласно рисунку 14 → *ОК*.

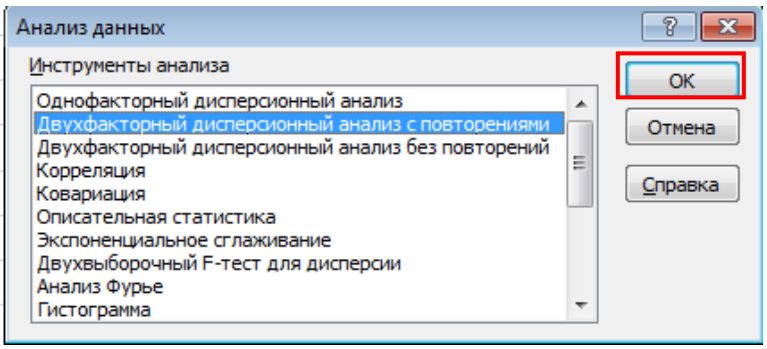


Рисунок 13 – Диалоговое окно *Анализ данных* в ППП MS EXCEL.

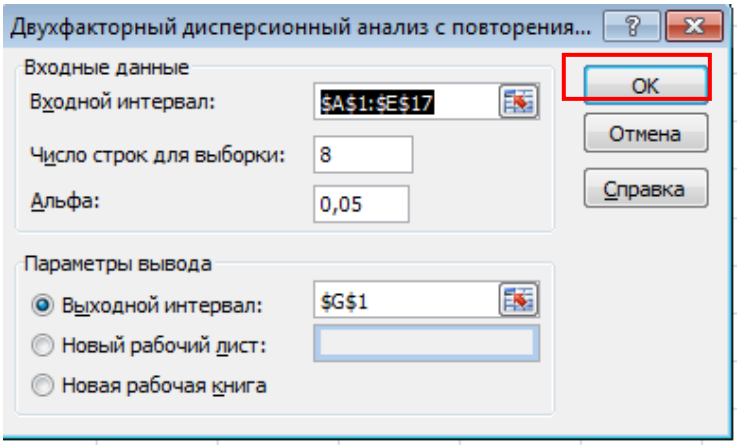


Рисунок 14 – Диалоговое окно *Двухфакторный дисперсионный анализ с повторениями* в ППП MS EXCEL.

Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Выборка	70,14063	1	70,14063	0,885929	0,350625721	4,012973319
Столбцы	1337,047	3	445,6823	5,629301	0,001914998	2,769430949
Взаимодействие	58,67188	3	19,55729	0,247023	0,8631111	2,769430949
Внутри	4433,625	56	79,17188			
Итого	5899,484	63				

Рисунок 15 – Результаты Двухфакторного дисперсионного анализа в ППП MS EXCEL.

11. Результаты двухфакторного дисперсионного анализа представлены на рисунке 15. Столбец *P-Значение* отражает информацию для принятия решений:

Значение 0,35 в строке *Выборка* превышает заданный уровень значимости $\alpha = 0{,}05$, следовательно, нет оснований отвергнуть нулевую гипотезу H_0^B . Таким образом, **пол сотрудника не оказывает влияние на среднюю производительность труда.**

Значение 0,0019 в строке *Столбцы* не превышает заданный уровень значимости $\alpha = 0{,}05$, следовательно, гипотеза H_0^A отвергается. Таким образом, **стаж работы оказывает влияние на среднюю производительность труда.**

Значение 0,86 в строке *Взаимодействие* превышает заданный уровень значимости $\alpha = 0{,}05$, следовательно, гипотеза H_0^{AB} не отвергается. Таким образом, **совместное влияние стажа работы и пола работника на среднюю производительность труда отсутствует.**

12. Построим графики средних значений в разрезе уровней факторного признака. Для этого используем промежуточные итоги результатов двухфакторного дисперсионного анализа в ППП MS EXCEL (рисунок 16). Перенесём данные для построения графиков на новый Лист MS EXCEL (рисунок 17).

G	H	I	J	K	L
Двухфакторный дисперсионный анализ с повторениями					
ИТОГИ	до 5	5-10	10-15	15-20	Итого
м					
Счет	8	8	8	8	32
Сумма	1258	1276	1281	1360	5175
Среднее	157,25	159,5	160,125	170	161,71875
Дисперсия	76,21429	87,42857	73,26786	60,85714	92,01512097
ж					
Счет	8	8	8	8	32
Сумма	1279	1278	1321	1364	5242
Среднее	159,875	159,75	165,125	170,5	163,8125
Дисперсия	121,5536	62,5	65,55357	86	96,02822581

Рисунок 16 – Результаты Двухфакторного дисперсионного анализа в ППП MS EXCEL.

	A	B	C	D	E
1		до 5	5-10	10-15	15-20
2	м	157,25	159,5	160,125	170
3	ж	159,875	159,75	165,125	170,5
4					
5					
6		м	ж		
7	до 5	157,25	159,875		
8	5-10	159,5	159,75		
9	10-15	160,125	165,125		

Рисунок 17 – Исходные данные для построения графиков средних в ППП MS EXCEL.

На панели меню выбираем *Вставка* → *График*. В появившемся окне щёлкаем правой кнопкой мыши. Из раскрывающегося списка выбираем *Выбрать данные*. Диалоговое окно *Выбор источника данных* заполняем согласно рисунку 18 → *ОК*. Средняя производительность труда в разрезе стажа работы представлена на рисунке 19.

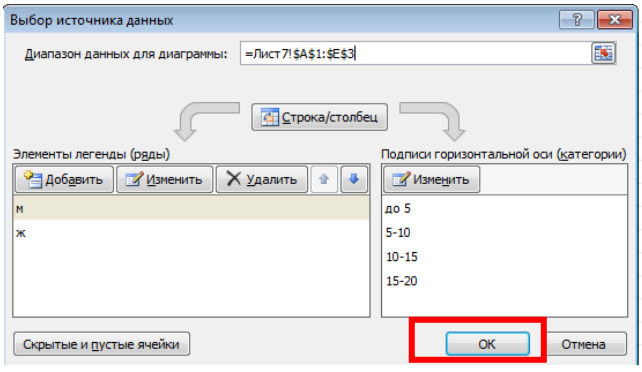


Рисунок 18 – Диалоговое окно *Выбор источника данных* в ППП MS EXCEL.

Анализ графиков на рисунке 19 позволяет сделать выводы о том, что средняя производительность по мужчинам и женщинам практически не отличается в категории стажа 5 – 10 лет и 15 – 20 лет. В категориях стажа до 5 лет и 10 – 15 лет средняя производительность по женщинам отличается примерно на 3 детали в день и на 5 деталей в день соответственно.

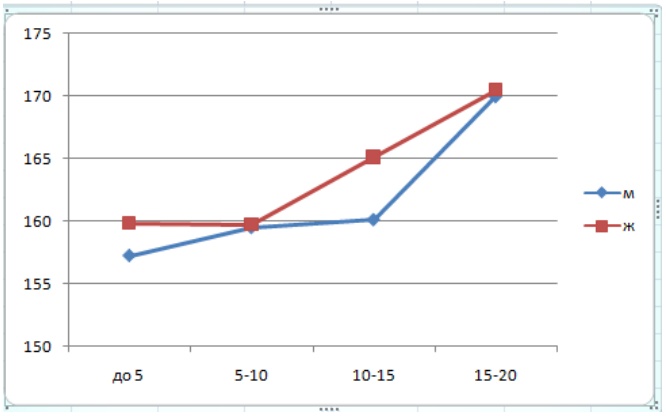


Рисунок 19 – Средняя производительность труда в разрезе стажа работы.

Для построения графиков средних в разрезе пола сотрудника снова выбираем *Вставка* → *График*. В появившемся окне щёлкаем правой кнопкой мыши. Из раскрывающегося списка выбираем *Выбрать данные*. Диалоговое окно *Выбор источника данных* заполняем согласно рисунку 20 → *ОК*. Средняя производительность труда в разрезе стажа работы представлена на рисунке 21.

Выводы по графикам рисунка 21 дублируют ранее сформулированные выводы. Средняя производительность труда практически не отличается для мужчин и женщин в категории стажа 5 – 10 лет и 15 – 20 лет. Наибольший скачок средней производительности в сторону женщин наблюдается в категории стажа 10 – 15 лет.

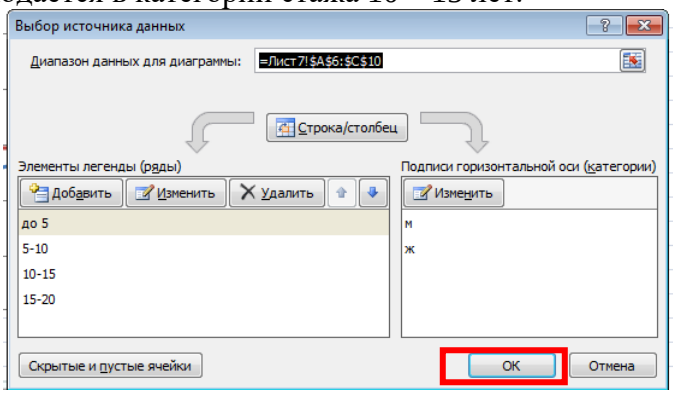


Рисунок 20 – Диалоговое окно *Выбор источника данных* в ППП MS EXCEL.

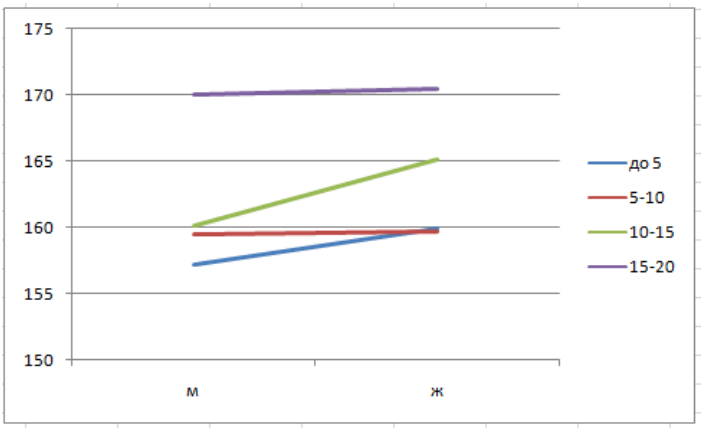


Рисунок 21 – Средняя производительность труда в разрезе пола сотрудника

Задание:

1. Согласно порядковому номеру в журнале старосты выбрать вариант задачи.
2. Сформулировать нулевые гипотезы для проведения однофакторного и двухфакторного дисперсионного анализа при уровне значимости α=5%.
3. Проверить нулевые гипотезы с помощью ППП MS EXCEL. Получить графическое представление средних значений в разрезе уровней факторного признака.
4. Сформулировать ключевые выводы.
5. Спроектировать модель машинного обучения по кластеризации поло-возрастных групп населения и их производительности труда с использованием дисперсионного анализа (применительно к кластеризации).

Отчёт по практическому занятию должен содержать:

— постановку задачи (в том числе исходные данные, цель занятия);

— нулевые гипотезы для проведения однофакторного и двухфакторного

— дисперсионного анализа;

— скриншоты таблиц однофакторного и двухфакторного дисперсионных

— анализов, полученных с помощью ППП MS EXCEL с выводами по принятию/ отклонению нулевых гипотез;

— графическое представление средних значений в разрезе уровней факторного признака, экономические выводы из графиков;

— сформированную теоретическую модель машинного обучения, согласно задания практического занятия;

— общий вывод по занятию, в том числе итог относительно достижения цели практического занятия.

Типовые оценочные материалы по Теме 7. «Распределенная обработка информации. Организация интернет- интернет- систем научных и производственных программ» и Теме 8. «Применение средств сетевых информационных технологий телекоммуникаций в науке с интеграцией систем машинного обучения для анализа информации»

«Машинное обучение в сетевых информационных технологиях телекоммуникаций»: Настройка данных для анализа и проектирование модели данных разговорного чат-бота (ассистента) для финансовой организации

Цель занятия: рассмотреть понятие, определение, формы и методы машинного обучения, используя описание выполнения типового задания по разработке и моделированию разговорного интерфейса (чат-бота) согласно предложенных вариантов, результирующим заданием – отчетом, разработать чат-бот для целей финансовой организации.

Ход работы:

1. Изучить справочный материал
2. Разработать и протестировать чат-бота на DialogFlow, умеющего общаться на тему вашего варианта.
3. Сформировать отчет в MS Word в виде скриншотов экрана. Вставить в отчет ссылку на бота.

Варианты

№ варианта	Тема	№ варианта	Тема
1	Интеллектуальная поддержка управленческих решений. Экспертные системы	9	Назначение и функциональные возможности табличного процессора MS Excel

№ варианта	Тема	№ варианта	Тема
2	Понятие модели данных	10	Цели создания информационных систем, состав информационной системы
3	Основные сервисы сети Интернет	11	Определение и классификация информационных технологий
4	Классификация компьютерных сетей	12	Информационно-поисковые системы (машины) сети Интернет
5	Варианты реализации сетевых топологий компьютерных сетей	13	Классификация основных методов защиты информации
6	Виды памяти персонального компьютера	14	Интеллектуальные информационные системы: определение и области применения
7	Состав вычислительной системы. Принципы фон Неймана	15	Определение и структура базы данных
8	Кодирование информации	16	Экспертные системы. Краткая характеристика. Общие задачи

Справочный материал:

Для тестового примера создадим бота, который отвечает на типовые вопросы.

1. Зайдите на сайт https://dialogflow.com/?ref=vc.ru

2. Нажмите кнопку Sign in для авторизации. Если у вас не авторизован аккаунт, то Google попросит это сделать.

3. Разрешите приложению Dialogflow использовать наш Google-аккаунт. Мы остаёмся внутри инфраструктуры Google, поэтому не беспокойтесь о безопасности.

4. После авторизации нажмите Go to console.

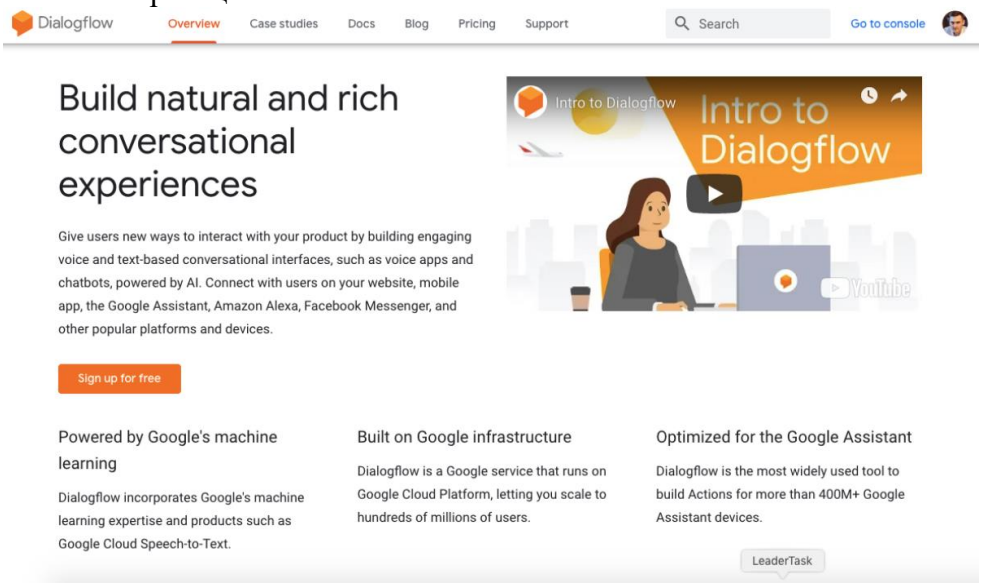


Рис. Первый экран Dialogflow

5. Система попросит авторизоваться ещё раз и согласовать необходимые разрешения.



6. В итоге попадете в раздел, где нужно подтвердить соглашение, делать это нужно только при запуске первого бота, в дальнейшем процесс будет значительно короче.

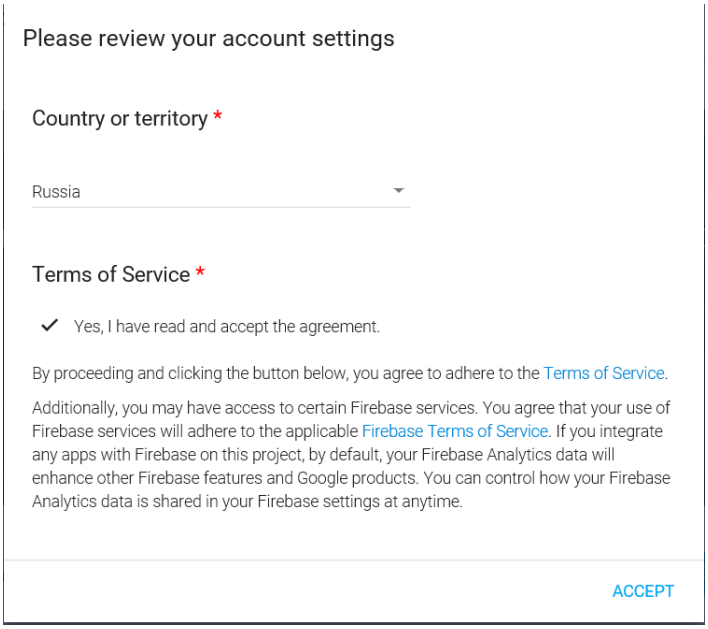


Рис. Подтверждение вашего согласия на использование сервиса

7. Далее откроется окно, где уже можно создать своего первого бота.

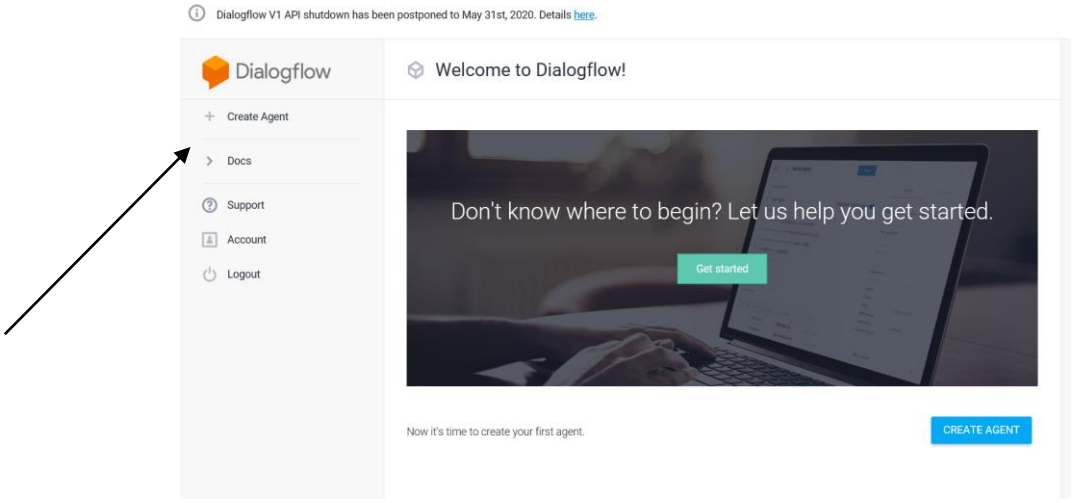


Рис. Общий вид Dialogflow до запуска первого бота

8. Нажмите Create agent.

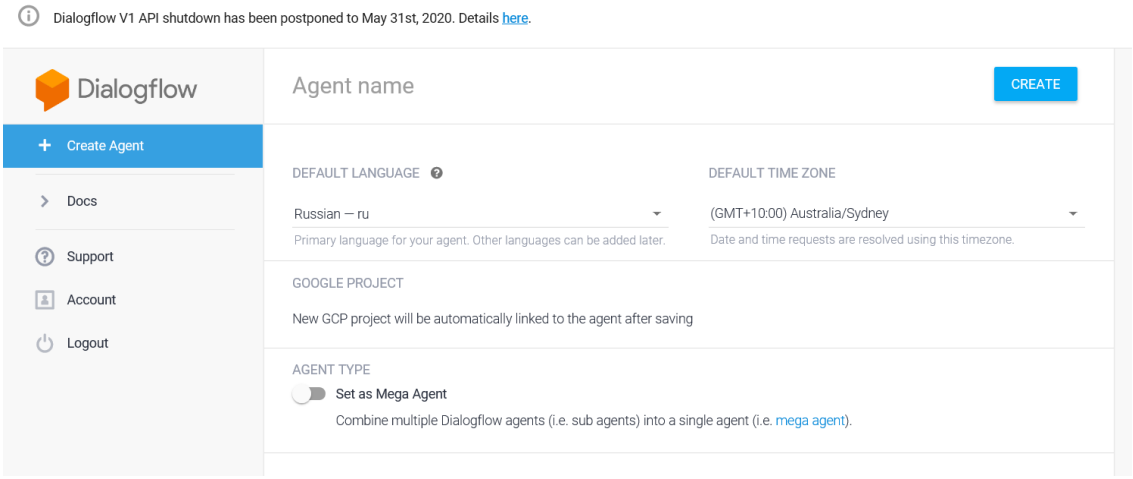


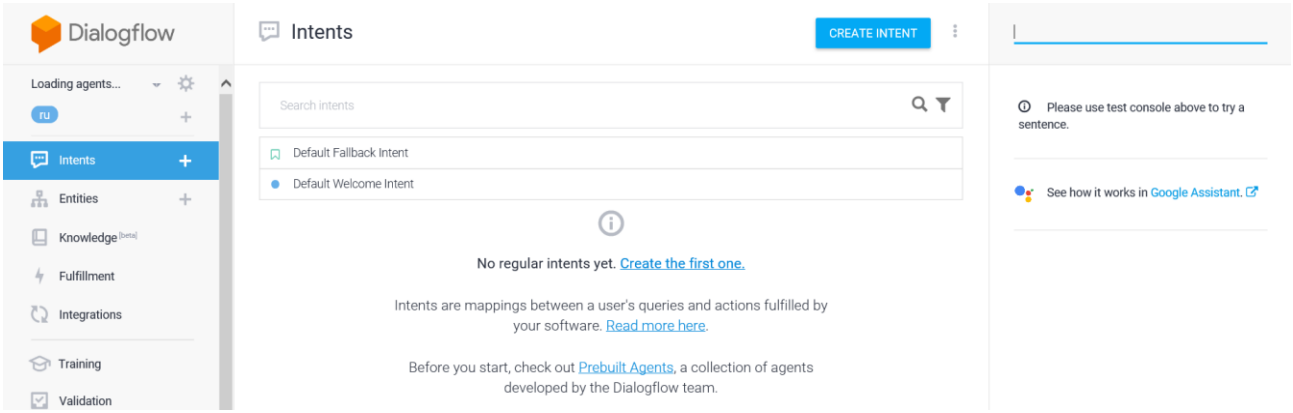
Рис. Экран создания бота

9. Придумайте ему название. Заполнить поле «Имя» латинскими буквами (Agent name)

10. Выберать русский язык (поле Default Language), если бот будет говорить на русском. Dialogflow постоянно обучается и распознаёт текст с учётом особенностей, поэтому правильное указание языка повысит точность распознавания. Если вы забудете это сделать, бот всё равно будет работать, только стандартные ответы будут на английском, а не на русском, и вам придётся их редактировать. Поэтому лучше выбрать сразу нужный язык.

11. Нажмите Create. Бот создан.

12. На экране отражено, уже создано 2 intent: Default Fallback и Default Welcome. Два созданных по умолчанию intent-а



В этот момент бот уже умеет отвечать на русском на «Привет» и другие стандартные приветствия и говорить, что он нас не понимает, если ни один intent не подошёл (в нашем случае пока мы либо отвечаем на «привет», либо отвечаем, что не понимаем собеседника).

13. Для простоты тестирования бота используем веб-версию. Для этого выбираем раздел Integrations слева окна и там выбираем первый блок.

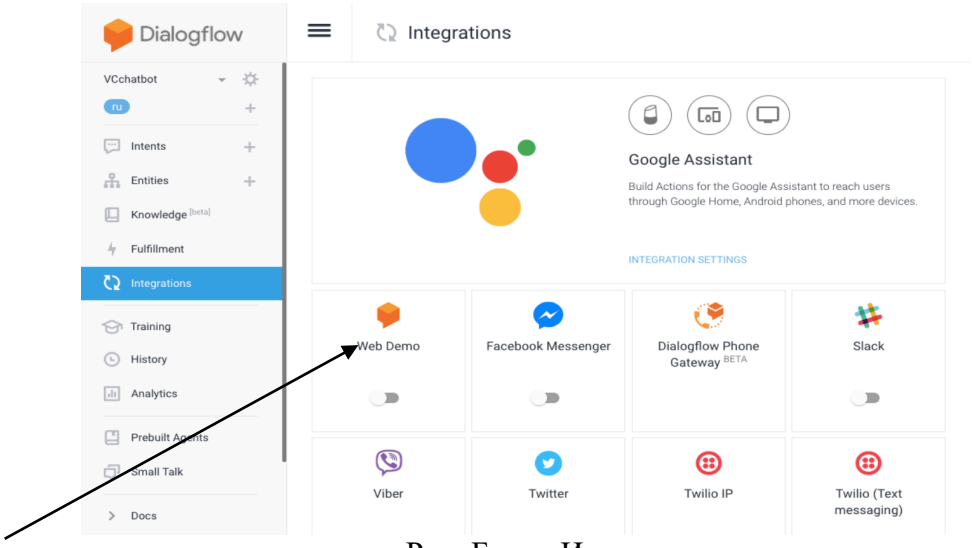


Рис. Блок «Интеграции»

14. Отредактировать текущие intent под Вашу модель распознавания фраз. В целом мы можем просто воспринимать эти сущности как вопросы-ответы. Да, при усложнении можно между ними настраивать последовательности и взаимосвязи, распознавания и интеграции со внешними сервисами.

Редактирование стандартных intent.

15. Редактирование с intent, который выдаёт информацию, если бот не понимает собеседника, – Default Fallback Intent. Для этого в левом меню нужно вернуться на пункт Intents и нажать на Default Fallback Intent.

16. В настройках этого блока нас интересует только раздел Responses. Удалить там лишнее, добавьте свои строчки. Система будет выдавать в случайном порядке один из ответов. Поэтому можно дополнить, оставить один вариант.

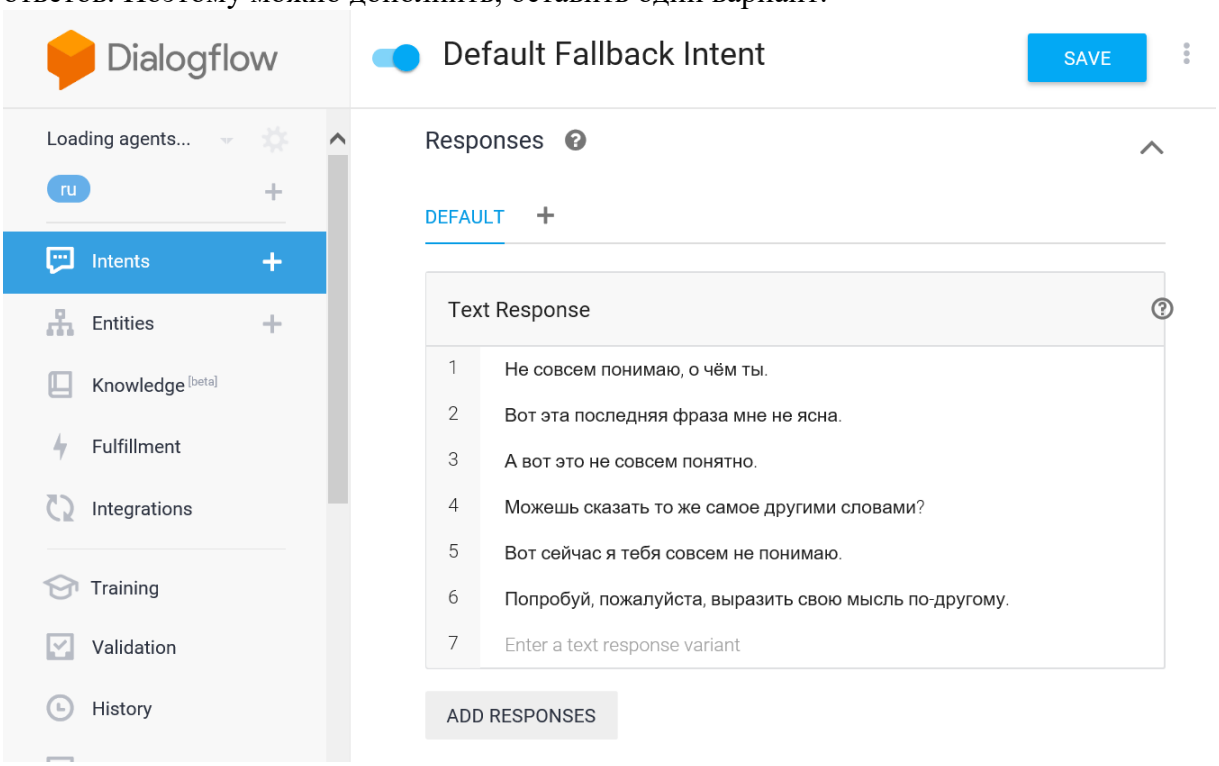


Рис. Default Fallback Intent по умолчанию

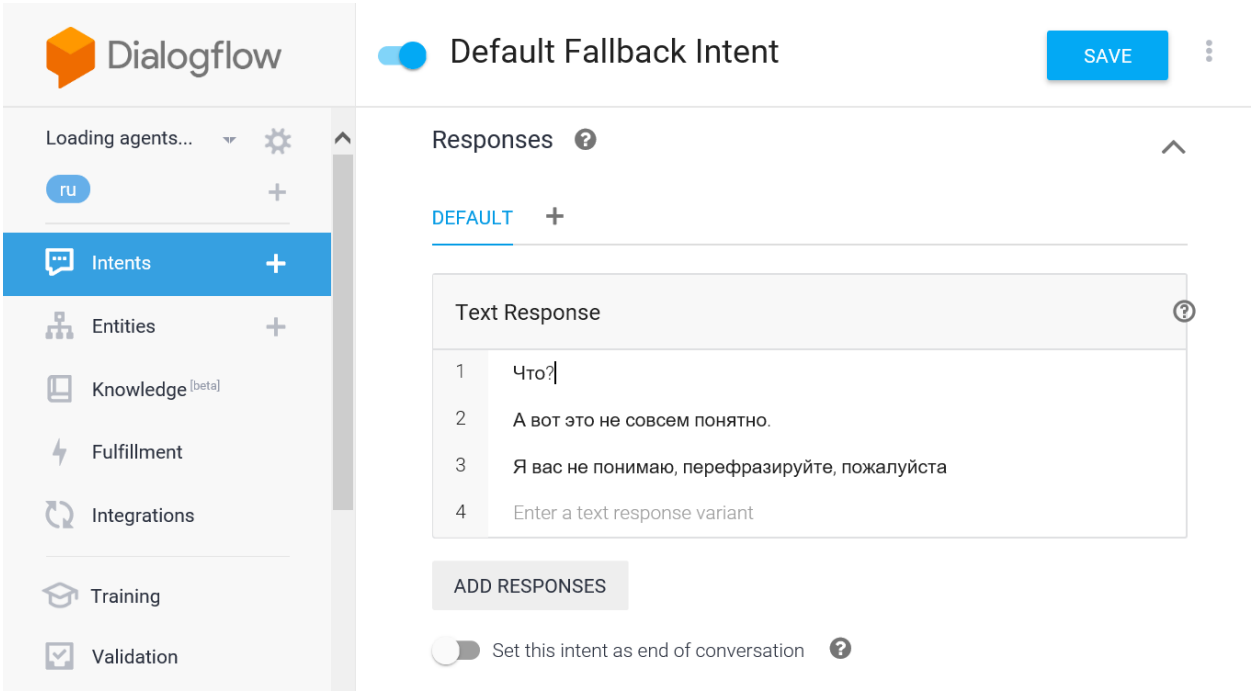


Рис. Default Fallback Intent после изменений

17. Всё это сразу можно проверить в тестовой веб-версии.

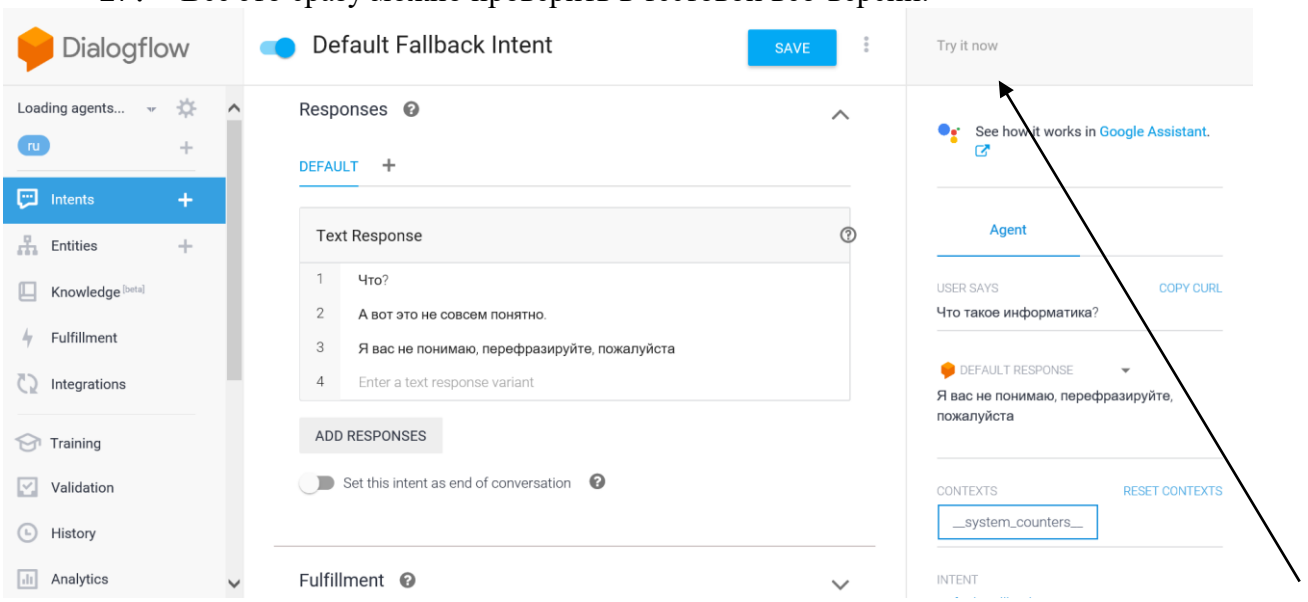


Рис. Проверка ответа бота

18. Нажмите после изменения на кнопку Save.

19. Перейдём к редактированию intent Welcome. С ответами здесь логика такая же, поэтому описываем все варианты в Response, которые должен отправлять наш бот.

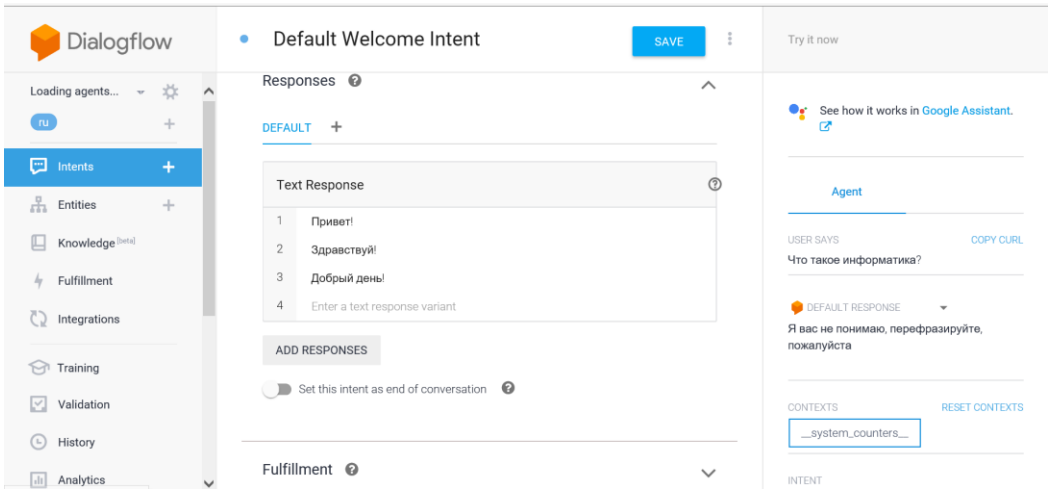


Рис. Default Welcome Intent после изменений

20. Кроме Response необходимо отредактировать блок Training phrases, который показывает, на какие слова должен реагировать этот intent. Отредактируйте этот блок по своему усмотрению.

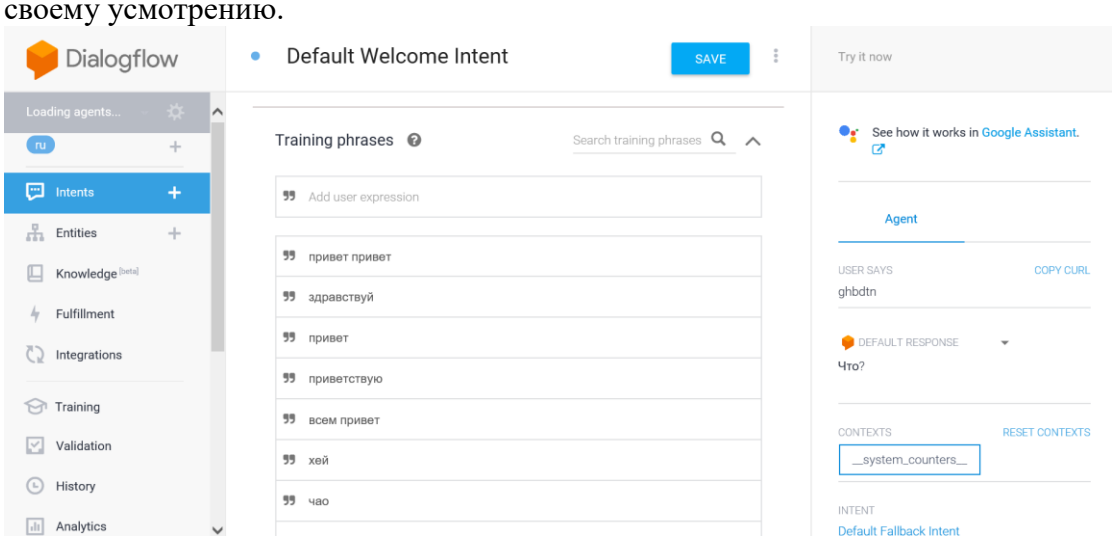


Рис. Default Welcome Intent

Создание новых intent

21. В разделе intents нажмите Create intent.

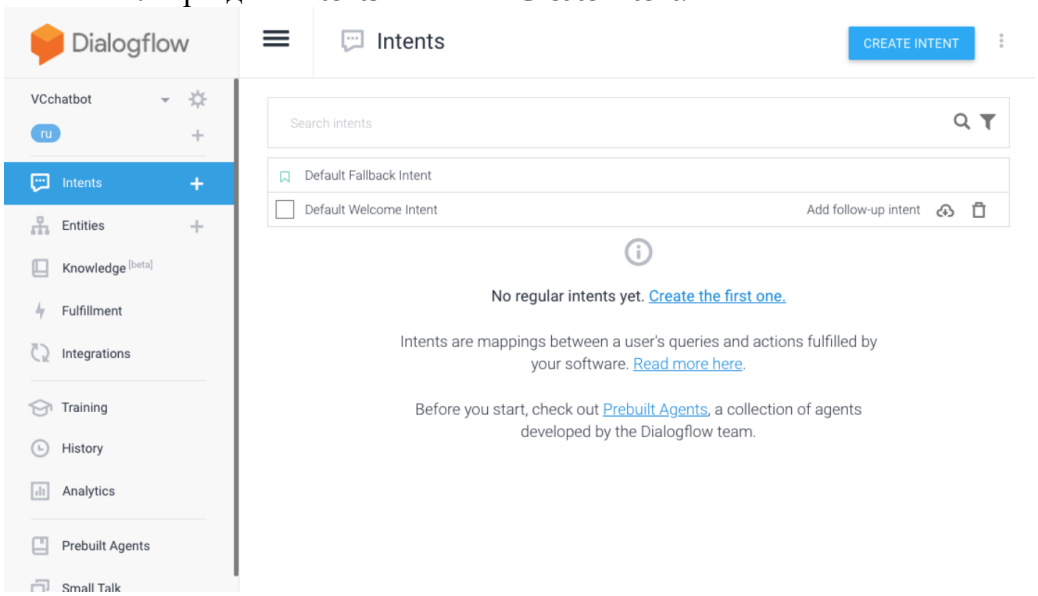


Рис. Create Intent

22. Заполнить поля «Название» и Training phrases (Add training phrases).

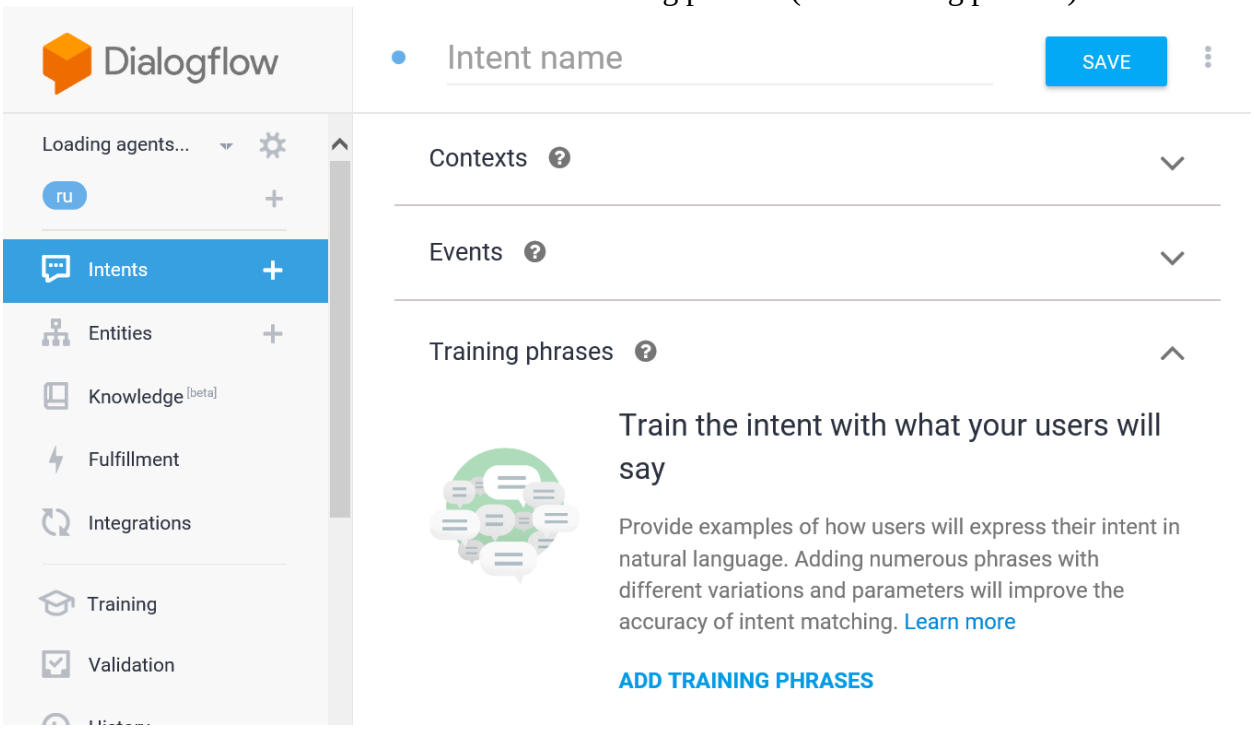
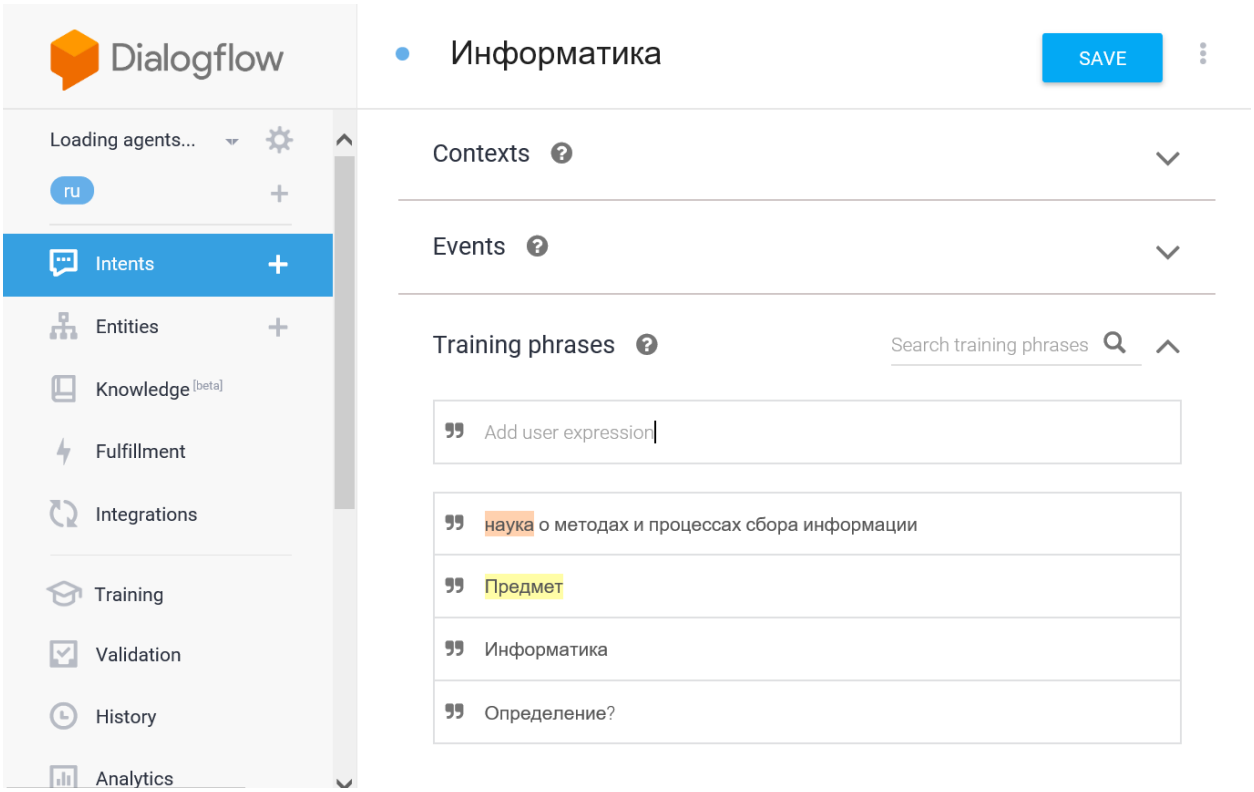


Рис. Add training phrases



23. А затем ответы — responses (Add responses).

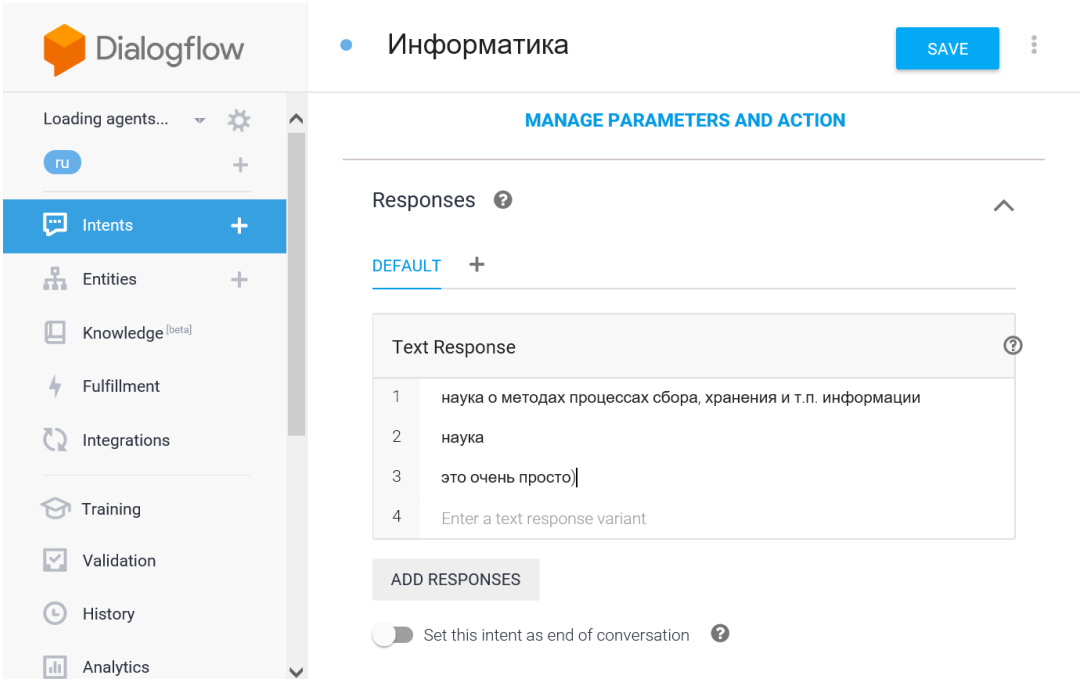
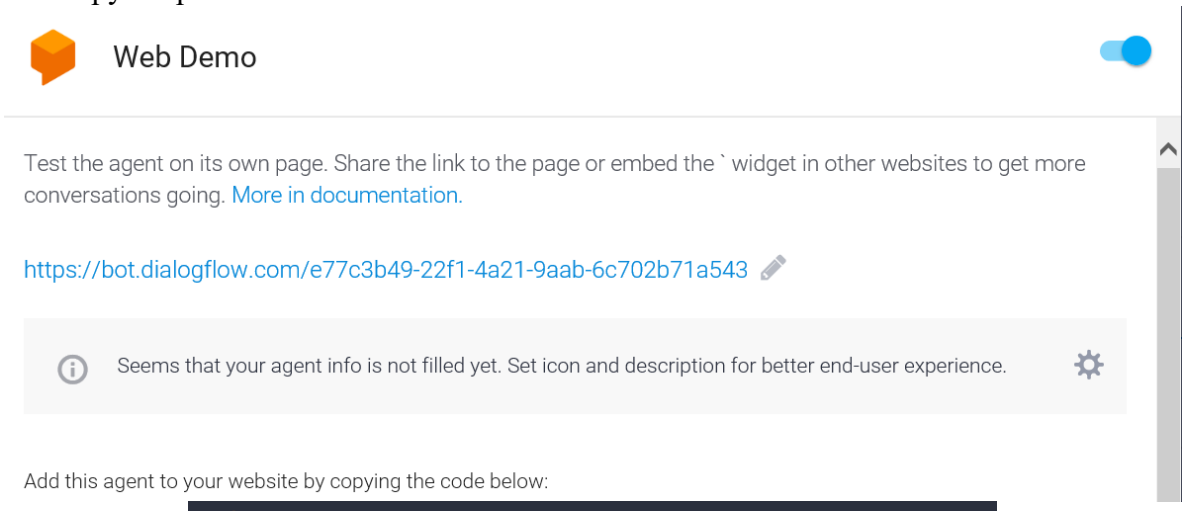


Рис. Add responses

24. Нажать Save и можем проверять работу в веб-версии. Перейдите по ссылке, которая автоматически сформировалась, когда вы переходили на веб-версию и протестируйте робота.



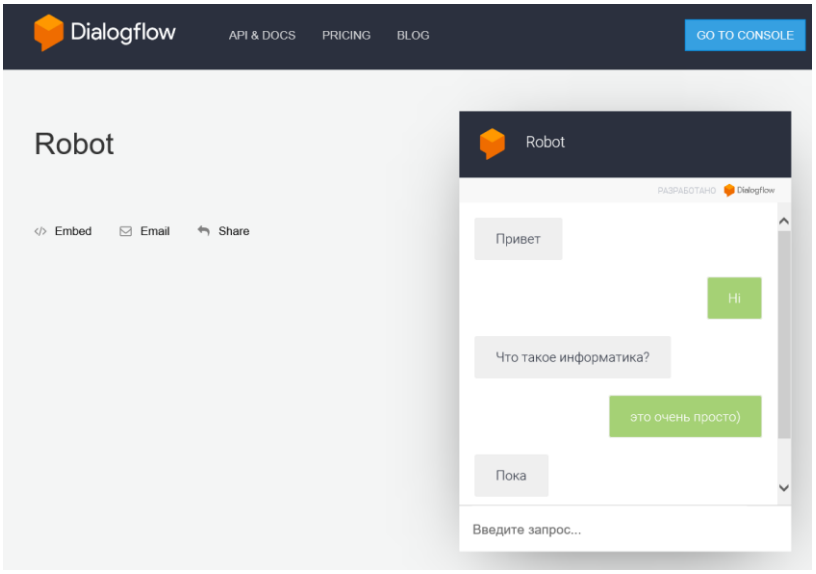


Рис. Пример работы тестового бота

25. Можно создать любое количество intent для решения Вашей задачи.

Интеграции

Dialogflow даёт множество возможных интеграций. Запустить его в Telegram предельно просто, использовать его с Facebook чуть сложнее, потому что нужно будет пройти модерацию самой соцсети, и это занимает некоторое время.

Отчёт по практическому занятию должен содержать:

— постановку задачи (в том числе исходные данные, цель занятия);

— нулевые гипотезы для построения разговорного интерфейса;

— описание данных обучения (эталонных данных) для разработанной модели машинного обучения;

— имя разговорного чат-бота в Telegram с кратким описанием тематики общения, и типовых вопросов / ответов.

— общий вывод по занятию, в том числе итог относительно достижения цели практического занятия.

Критерии качества оценки текущей аттестации

Критерии оценивания отчётов по практическим занятиям

5 баллов выставляется студенту, если он предоставил письменный подробно изложенный отчёт с описанием всех его манипуляций при выполнении практического задания, провёл необходимое тестирование результатов, полученных в конце выполнения практического занятия, с использованием встроенных механизмом программного обеспечения проверил результаты расчётов на отсутствие ошибок в них.

4 балла выставляется студенту, если он предоставил письменный, но не подробный отчёт с описанием большинства его манипуляций (но не всех) при выполнении практического задания, провёл необходимое тестирование результатов, полученных в конце выполнения практического занятия, с использованием встроенных механизмом программного обеспечения проверил результаты расчётов на отсутствие ошибок в них.

3 балла выставляется студенту, если он предоставил письменный, но не подробный отчёт с описанием большинства его манипуляций (но не всех) при выполнении практического задания.

2 балла - выставляется студенту, если он сдал письменный отчёт не полным и без проведения необходимых проверок расчётов на ошибки, при этом описание действий в отчёте уже содержит структурные ошибки, при условии, что выводы оформлены верно.

0 баллов - выставляется студенту, если он не сдал письменный отчёт.

Критерии оценивания устного опроса

Опрос/опорный конспект по теме оценивается по следующим критериям:

1 балл – обучающийся верно, самостоятельно отвечает на поставленный вопрос,

0 балл – обучающийся испытывает затруднения при ответе на поставленный вопрос или не может ответить на поставленный вопрос.

Шкала оценивания результатов успеваемости при осуществлении текущего контроля

Вид оценочного средства	Шкала оценивания (баллы)
Опрос/опорный конспект	0-6
Практическое занятие (предусмотрены шесть практических занятий)	0-30
Тестирование	0-10
Аналитическое эссе (предусмотрено три эссе)	0-24
Итого	***0-70****

* 70 - максимальное количество баллов по результатам текущего контроля

5. Оценочные материалы для промежуточной аттестации

5.1. Экзамен проводится с применением следующих средств: в устной и письменной форме: в устной форме по заданиям закрытых экзаменационных билетов с теоретическими вопросами и расчетными практическими заданиями, в письменной форме с применением компьютерного тестирования.

Процедура проведения экзамена установлена в соответствии с Регламентом по проведению промежуточной аттестации с применением балльно-рейтинговой системы оценки знаний, обучающихся в Дальневосточном институте управления – филиале РАНХиГС.

5.2. Оценочные материалы промежуточной аттестации

Компонент компетенции	Промежуточный/ключевой индикатор оценивания	Критерий оценивания
ОПК ОС-3.1	использует современные инструментальные методы экономического анализа для решения исследовательских и практических задач / самостоятельно использует современные инструментальные методы экономического анализа для решения исследовательских задач	В типовых условиях демонстрирует знание свойств и ограничений выбранной или заданной расчетной эконометрической модели; умеет определять источники информации для решения исследовательских задач. Производит самостоятельный расчёт показателей экономических моделей с применением современных инструментальных методов экономического анализа.
ОПК ОС-5.1	использует современные информационные технологии и программные средства для решения профессиональных задач / самостоятельно использует современные информационные технологии и программные средства для решения профессиональных задач	В типовых условиях демонстрирует знание о современных информационных технологиях и программных средств для решения профессиональных задач. Умеет использовать специализированные программы для оценки параметров модели и проверки адекватности модели. Самостоятельно использует современные информационные технологии и программные средства для решения профессиональных задач.

Типовые оценочные материалы промежуточной аттестации

Типовые вопросы для подготовки к экзамену

1. Линейная регрессия или метод случайного леса
2. Дисперсия
3. Регуляризация
4. Градиентный спуск
5. Интерпретация коэффициенты линейной регрессии
6. Метод максимального правдоподобие
7. Различие между методами классификации и регрессии в машинном обучении
8. Линейный дискриминантный анализ
9. Алгоритмы бэггинг и бустинга.
10. Встраивание слов GloVe (глобальные векторы для представления слов).
11. Разница между Лассо и регрессией гребня
12. Методы предотвращения переобучения

13. Перекрёстная проверка,
14. Регуляризация (математика)
15. Вербализация нейронных сетей
16. Априорная вероятность
17. Байесовское сравнение моделей (англ. *bayesian model comparison*),
18. Использование MSE.
19. Взаимосвязь между PCA и LDA
20. Контролируемое и неконтролируемое обучением
21. размеченные данные
22. Частично контролируемое обучение
23. Алгоритм опорных векторов
24. Метод опорных векторов
25. Интегрированные программные пакеты. Интернет- технологии обработки образовательной информации на основе табличных процессоров, использования систем управления базами данных, интегрированных программных пакетов.
26. Особенности современных технологий решения задач текстовой и графической обработки, табличной и математической обработки, накопления и хранения данных с помощью систем машинного обучения.
27. Представление знаний. Базы знаний (БЗ). Экспертные системы (ЭС). Примеры экспертных систем соответствующей научной области. Интеллектуальные интернет-системы (ИИС).
28. Методы поиска информации в современных системах с помощью систем машинного обучения.
29. Онлайновые библиотечные каталоги: ИНИОН; РГБ; ВГБИЛ; Юридической Научной Библиотеки; университетских библиотек. Книжные интернет-магазины как средство поиска научной информации. Электронные интернет- ресурсы: классификация, производители, общие правила работы.
30. Интернет- технологии образовательных программ. Технология применения электронных образовательных программ. Интернет- технологии документационного обеспечения образовательных программ.
31. Проблемы технологий в учебном процессе. Теоретико-методологические основы технологизации процесса обучения. Образовательные и обучающие технологии на современном этапе. Проблемы и перспективы информатизации высшего образования.
32. Интернет- системы в деятельности палат Федерального Собрания Российской Федерации. Автоматизированная система обеспечения законодательной деятельности, ее назначение и функции.
33. Картотека судебных дел как интегрированный информационный ресурс арбитражных судов Российской Федерации.
34. Стратегические ориентиры и актуальные задачи развития электронного правительства в России.
35. Интернет – основа единого международного информационного пространства. Образовательные ресурсы и услуги Интернета. Поиск информации в Интернете. Интернет- ресурсы: WWW сервер, файловый сервер, сервер телеконференций.
36. Принципы организации глобальных сетей. Специальные методы поиска информации в мировой глобальной сети Интернет.
37. Дистанционное обучение (ДО) и открытое образование. Дистанционные образовательные технологии (ДОТ). Основные принципы дистанционного обучения: гибкость в выборе места и времени обучения, возможность обучения без отрыва от основной деятельности, доступность для живущих в отдаленной местности, свобода выбора учебного заведения и дисциплин для обучения.
38. Решение проблем обеспечения возможности для получения высшего образования и обучения на протяжении всей жизни, предоставления учащимся права

свободного выбора места, времени и технологий обучения в рамках системы открытого образования с помощью систем машинного обучения.

39. Особенности применения интернет- телекоммуникаций в образовании. Образовательные услуги сети Интернет. Ресурсы Интернет. Дидактические свойства Интернет.

40. Дидактические функции интернет- телекоммуникаций. Организация дистанционного обучения с помощью систем машинного обучения.

41. Разработка электронных учебно-методических комплексов. Технологии компьютерного тестирования, обработки и интерпретации результатов тестов с помощью систем машинного обучения.

42. Специализированные Интернет-сайты как инструмент методической поддержки учебного процесса.

Шкала оценивания результатов обучения при осуществлении промежуточной аттестации

Критерий оценивания	Оценка
В типовых условиях демонстрирует знание закономерностей и методов экономической науки, понимает свойства и ограничения выбранной или заданной расчетной эконометрической модели; умеет определять источники информации для проведения эконометрического анализа экономического объекта. Производит самостоятельный расчёт показателей экономических моделей с применением современных инструментальных методов экономического анализа и современных информационных технологий.	Отлично
В типовых условиях с небольшими ошибками демонстрирует знание закономерностей и методов экономической науки, понимает свойства и ограничения выбранной или заданной расчетной эконометрической модели; умеет определять источники информации для проведения эконометрического анализа экономического объекта. Производит самостоятельный расчёт показателей экономических моделей с применением современных инструментальных методов экономического анализа и современных информационных технологий без критических (фатальных) ошибок.	Хорошо
В типовых условиях частично верно демонстрирует знание закономерностей и методов экономической науки, частично верно понимает свойства и ограничения выбранной или заданной расчетной эконометрической модели. Частично верно умеет определять источники информации для проведения эконометрического анализа экономического объекта. Частично, верно, производит самостоятельный расчёт показателей экономических моделей с применением современных инструментальных методов экономического анализа и современных информационных технологий без критических (фатальных) ошибок.	Удовлетворительно
В типовых условиях не способен продемонстрировать знания о закономерностях и методах экономической науки, не понимает свойства и ограничения выбранной или заданной расчетной эконометрической модели; не умеет определять источники информации для проведения эконометрического анализа экономического объекта. Не способен производить расчёт показателей экономических моделей с применением современных инструментальных методов экономического анализа и современных информационных технологий. Или в расчётах допускает критические (фатальные) ошибки.	Не удовлетворительно

Шкала соответствия между баллами (процентами), выставляемыми по многобалльной (процентной) системе, и отметками бинарной системы оценки знаний

Наименование отметки	Сумма баллов (процентов)
«отлично»	100 – 90
«хорошо»	89 – 75
«удовлетворительно»	60 -74
«не удовлетворительно»	0-59

Промежуточный контроль – 30 баллов.

6. Методические материалы по освоению дисциплины

Методические указания для обучающихся по освоению материала занятий лекционного типа

Лекция представляет собой систематизированное, последовательное изложение преподавателем учебного материала в соответствии с изучаемой темой дисциплины. Занятия лекционного типа дают обучающимся систематизированные знания о наиболее сложных и актуальных проблемах изучаемой дисциплины.

Осуществляя учебные действия на лекционных занятиях, обучающиеся должны внимательно воспринимать материал, подготовленный преподавателем, мыслить, добиваться понимания изучаемого предмета. Обучающимся рекомендуется вести конспект. В случае недопонимания какой-либо части предмета следует в установленном порядке задать вопрос преподавателю.

Для визуального восприятия формулировки темы, учебных вопросов, наиболее важных определений, схем, иллюстраций, таблиц, графиков, дополнительных источников информации и т. п. обучающимся рекомендуется обращать внимание на предлагаемые преподавателем учебные презентации в формате MS PowerPoint (или созданные с помощью иных приложений).

Методические указания для обучающихся по освоению материала занятий семинарского типа

Занятия семинарского типа (практические занятия) являются аудиторными. Для того, чтобы аудиторный материал занятия был освоен обучающимся, ему необходимо предварительно изучить теоретический материал по данной теме, ознакомиться с литературой, рекомендованной преподавателем и найденной самостоятельно; выполнить задания, предложенные преподавателем к занятию; составить перечень вопросов, вызывающих затруднения, неясности или сомнения, обсудить их с преподавателем на занятии.

В ходе подготовки к семинарским занятиям обучающемуся рекомендуется изучить основную литературу, ознакомиться с дополнительной литературой, новыми публикациями в периодических изданиях. При этом следует учесть рекомендации преподавателя и требования рабочей программы. Рекомендуется также дорабатывать свой конспект лекции, делая в нем соответствующие записи из литературы, рекомендованной преподавателем и предусмотренной рабочей программой дисциплины.

Следует подготовить тезисы для выступлений по всем учебным вопросам, выносимым на практическое занятие.

Готовясь к выступлению на занятии, обучающийся может обращаться за методической помощью к преподавателю. Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам.

Методические указания для обучающихся по самостоятельной работе

Самостоятельная работа обучающихся направлена на самостоятельное изучение отдельных вопросов учебной дисциплины. Самостоятельная работа является обязательной для каждого обучающегося, ее объем определяется учебным планом.

Усвоение материала дисциплины в процессе аудиторных занятий и в результате самостоятельной подготовки и изучения отдельных вопросов дисциплины позволят студенту подойти к промежуточной аттестации подготовленным и потребует лишь повторения ранее пройденного материала. Знания, накапливаемые постепенно и в различных ракурсах, с использованием противоположных мнений и взглядов на ту или иную проблему экономики являются глубокими и качественными, и позволяют сформировать соответствующие компетенции как итог образовательного процесса.

Для систематизации знаний по дисциплине первоначальное внимание обучающемуся следует обратить на рабочую программу курса, которая включает основные темы дисциплины, в рамках которых и формируются вопросы для текущего контроля и промежуточной аттестации. Поэтому обучающийся, заранее ознакомившись с программой курса, может лучше сориентироваться в последовательности освоения курса с позиций организации самостоятельной работы.

При организации процесса освоения дисциплины обучающемуся следует учитывать:

1. Планирование времени, отведенного на освоение дисциплины.

При планировании времени на освоение дисциплины следует руководствоваться: а) структурой дисциплины (раздел 3 настоящей рабочей программы), в которой указаны количество академических часов в разрезе каждой темы, вида занятий (лекции, практические занятия) и часы на самостоятельную работу; б) формой текущего контроля успеваемости (опрос, опорный конспект, тестирование, реферат); в) формой промежуточной аттестации (зачет).

2. Последовательность действий при освоении дисциплины. Изучение каждой темы дисциплины целесообразно начинать со знакомства с содержанием дисциплины в разрезе тем; затем следует этап подбора источников из числа рекомендуемых и подобранных самостоятельно (научные статьи; информация с официальных сайтов государственных органов, др.). Изучение источниковой базы может сопровождаться конспектированием (например, тезисным), составлением глоссария, др. Целесообразно вести перечень проблемных вопросов как по существу темы, обусловленных пробелами в научном и правовом поле и проблемами практического характера, так и в случае затруднений с уяснением смысла изложенного в источниках материала (указанные вопросы могут быть разрешены впоследствии самим студентом во время занятий или на консультации с преподавателем).

3. Использование учебно-методических материалов и работа с литературой. Следует применять следующую последовательность подбора источников для изучения тем дисциплины: а) научная и учебная литература: основная и дополнительная; б) нормативные правовые материалы; в) Интернет-ресурсы и др.

4. Подготовка к текущему контролю успеваемости. При выполнении заданий следует руководствоваться методическими рекомендациями, представленными по каждому виду заданий (опрос/опорный конспект, реферат, тестирование).

При самостоятельной работе обучающиеся работают с рекомендованными материалами при минимальном участии преподавателя. При необходимости обучающемуся рекомендуется записывать вопросы, возникшие у него в ходе работы над темой, для обсуждения их с преподавателем в ходе семинаров или консультаций.

Методические указания для обучающихся по подготовке к опросу /составлению опорного конспекта

Опросы по темам проводятся в части проверки и уяснения основных экономических понятий и категорий. При подготовке к опросу обучающийся должен изучить вопросы для самостоятельной подготовки к практическим занятиям, рекомендуется составить краткий конспект ответа на каждый из представленных вопросов.

Конспект – это систематическая, логически связная запись материала, включающая в себя план изучаемого материала, выписки из учебника и дополнительных источников, тезисы материала, собственные мысли, выводы и т.п.

Методические указания по подготовке рефератов

Реферат – результат самостоятельной работы обучающегося, представляющий собой краткое изложение в письменном виде полученных результатов теоретического анализа определенной научной (учебно-исследовательской) темы, где автор раскрывает суть исследуемой проблемы, приводит различные точки зрения, а также собственные взгляды на нее. Реферат должен состоять из следующих частей: титульный лист; содержание; введение; основной текст, разбитый на главы и подразделы (вопросы); заключение; библиографический список. Реферат оформляется в соответствии с требованиями, предъявляемыми локальными нормативными актами РАНХиГС к письменным работам.

Эссе – очерк, трактующий какие-нибудь проблемы не в систематическом научном виде, а в свободной форме. Эссе – прозаическое сочинение небольшого объёма и свободной композиции, трактующее частную тему и представляющее попытку передать индивидуальные впечатления и соображения, так или иначе, с нею связанные. Эссе по дисциплине «Современные банковские технологии» – творческое мини-сочинение по конкретной проблеме, имеющей отношение к одной из тем дисциплины. Этапы написания эссе: 1) определение смысла предложенной проблемы, подбор и изучение научных статей, содержащих высказывания по ней; 2) осмысление проблемы в контексте дисциплины «Использование Интернет в экономической науке и производстве» (с какими проблемами налогообложения связана тема, что вы должны знать, чтобы её раскрыть); 3) формулирование своего отношения к высказыванию; 4) использование определений, терминов, обобщение, теоретическая обоснованность понятий и теорий; 5) использование фактов, примеров, обосновывающих вашу позицию; 6) вывод.

Методические рекомендации при применении ДОТ и регламент распределения видов работ по дисциплине с ДОТ

Данная дисциплина может быть реализована с применением элементов дистанционных образовательных технологий (ДОТ). Распределение видов учебной работы, форматов текущего контроля представлены в таблице:

Вид учебной работы	**Формат проведения**
Лекционные занятия	контактная аудиторная работа / в СДО
Практические занятия	контактная аудиторная работа / в СДО
Самостоятельная работа	контактная аудиторная работа / в СДО
Промежуточная аттестация	Устно /устно в ДОТ
Формы текущего контроля	**Формат проведения**
Опрос/опорный конспект	контактная аудиторная работа / в СДО
Отчёт по практическому занятию	контактная аудиторная работа / в СДО
Тестирование	контактная аудиторная работа / в СДО
Аналитическое эссе	контактная аудиторная работа / в СДО

Доступ к системе дистанционных образовательных технологий (СДО) осуществляется каждым обучающимся самостоятельно с любого устройства на образовательном портале института: https://lms.ranepa.ru. Пароль и логин к личному кабинету / профилю предоставляется студенту в деканате.

Все формы текущего контроля, проводимые в системе дистанционного обучения, оцениваются в системе дистанционного обучения. Доступ к видео и материалам лекций предоставляется в течение всего семестра. Доступ к каждому виду работ и количество попыток на выполнение задания предоставляется на ограниченное время согласно регламенту дисциплины, опубликованному в СДО. Преподаватель оценивает выполненные обучающимся работы после окончания срока выполнения согласно установленному графику.

Вопросы для самостоятельной подготовки к занятиям семинарского типа

1. Линейная регрессия или метод случайного леса
2. Дисперсия
3. Регуляризация
4. Градиентный спуск
5. Интерпретация коэффициенты линейной регрессии
6. Метод максимального правдоподобие
7. Различие между методами классификации и регрессии в машинном обучении
8. Линейный дискриминантный анализ
9. Алгоритмы бэггинг и бустинга.
10. Встраивание слов GloVe (глобальные векторы для представления слов).
11. Разница между Лассо и регрессией гребня
12. Методы предотвращения переобучения
13. Перекрёстная проверка,
14. Регуляризация (математика),
15. Вербализация нейронных сетей,
16. Априорная вероятность,
17. Байесовское сравнение моделей (англ. bayesian model comparison),
18. Использование MSE.
19. Взаимосвязь между PCA и LDA
20. Контролируемое и неконтролируемое обучением
21. Размеченные данные
22. Частично контролируемое обучение
23. Алгоритм опорных векторов
24. Метод опорных векторов
25. Метод классификатора с максимальным зазором.
26. Модель обнаружения банковского мошенничества
27. Эвристические алгоритмы
28. Модель отказа в выдаче кредита.
29. Коэффициенты логистической регрессии для категориальных и булевых переменных
30. Модель ML для прогнозирования уровня заполняемости отеля

7. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

7.1 Основная литература

1. Машинное обучение в финансах : учебник / С. Ю. Богатырев, А. А. Помулев, А. В. Затевахина [и др.] ; под. ред. С. Ю. Богатырева. - Москва : Прометей, 2024. - 224 с. - ISBN 978-5-00172-572-5. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2144368 . – Режим доступа: по подписке.

2. Гниденко, И. Г. Технологии и методы программирования : учебное пособие для вузов / И. Г. Гниденко, Ф. Ф. Павлов, Д. Ю. Федоров. — Москва : Издательство Юрайт, 2020. — 235 с. — (Высшее образование). — ISBN 978-5-534-02816-4. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/450999 Москва : Издательство Юрайт 2020 https://urait.ru/bcode/450999

3. Чио, К. Машинное обучение и безопасность: защита систем с помощью данных и алгоритмов : практическое руководство / К. Чио, Д. Фримэн. - Москва : ДМК Пресс, 2020. - 388 с. - ISBN 978-5-97060-713-8. - Текст : электронный. - URL: https://znanium.com/catalog/product/1908430 . – Режим доступа: по подписке.

7.2 Дополнительная литература

1. Баженов, Р. И. Интеллектуальные информационные технологии в управлении : учебное пособие / Р. И. Баженов. — Москва : Ай Пи Ар Медиа, 2023. — 124 с. — ISBN 978-5-4497-1864-8. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/127570.html (вход по паролю).

2. Информационные технологии в экономике и управлении в 2 ч. Часть 1 : учебник для вузов / В. В. Трофимов [и др.] ; под редакцией В. В. Трофимова. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2020. — 269 с. — (Высшее образование). — ISBN 978-5-534-09083-3. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/456061 Москва : Издательство Юрайт 2020 https://urait.ru/bcode/456061

3. Информационные технологии в экономике и управлении в 2 ч. Часть 2 : учебник для вузов / В. В. Трофимов [и др.] ; под редакцией В. В. Трофимова. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2020. — 245 с. — (Высшее образование). — ISBN 978-5-534-09084-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/456062 Москва : Издательство Юрайт 2020 https://urait.ru/bcode/456062

4. Нетёсова, О. Ю. Информационные системы и технологии в экономике : учебное пособие для вузов / О. Ю. Нетёсова. — 3-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 178 с. — (Высшее образование). — ISBN 978-5-534-08223-4. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/452595 Москва : Издательство Юрайт2020 https://urait.ru/bcode/452595

5. Суворова, Г. М. Адаптивные информационные и коммуникационные технологии в управлении средой обитания : учебное пособие для среднего профессионального образования / Г. М. Суворова. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 210 с. — (Профессиональное образование). — ISBN 978-5-534-15192-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/520366 (вход по паролю).

6. Черткова, Е. А. Компьютерные технологии обучения : учебник для вузов / Е. А. Черткова. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 250 с. — (Высшее образование). — ISBN 978-5-534-07491-8. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/452449 Москва : Издательство Юрайт 2020 https://urait.ru/bcode/452449 (вход по паролю).

7.3 Нормативно-правовые акты

1. Конституция Российской Федерации (с внесенными поправками) // В соответствии с Указом Президента от 3 июля 2020 года «Об официальном опубликовании

Конституции Российской Федерации с внесенными в нее поправками» текст размещен на «Официальном интернет-портале правовой информации» (pravo.gov.ru).

2. Бюджетный кодекс Российской Федерации. [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс» https://www.consultant.ru/document/cons_doc_LAW_19702/ .

3. Гражданский кодекс РФ. – Ч. I, II. [текст]. [Электронный ресурс] / Официальный сайт Президента России: http://www.kremlin.ru/acts/bank/7279 .

4. Налоговый кодекс Российской Федерации: в 2-х томах. [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс» https://www.consultant.ru/document/cons_doc_LAW_19671/

5. Об информации, информационных технологиях и о защите информации : Федеральный закон № 149-фз от 27 июня 2006 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

6. Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления : Федеральный закон № 8-фз от 9 февраля 2009 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

7. Об организации предоставления государственных и муниципальных услуг : Федеральный закон № 210-фз от 27 июля 2010 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

8. Об электронной подписи : Федеральный закон № 63-фз от 6 апреля 2011 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

9. О персональных данных : Федеральный закон № 152-фз от 27 июля 2006 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

10. Об инфраструктуре, обеспечивающей информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных услуг в электронной форме: постановление Правительства РФ от 8 июня 2011 г. № 451 [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

11. О федеральных государственных информационных системах, обеспечивающих предоставление в электронной форме государственных и муниципальных услуг (осуществление функций) : постановление Правительства РФ от 24 октября 2011 г. № 861 [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

12. О мерах по обеспечению перехода федеральных органов исполнительной власти и органов государственных внебюджетных фондов на межведомственное информационное взаимодействие в электронном виде: постановление Правительства РФ от 28 декабря 2011 г. № 1184 [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

13. О правилах организации наличного денежного обращения на территории Российской федерации: Положение Банка России № 14-П от 05.01.98 г. – Режим доступа: http://www.cbr.ru, свободный. – Загл. с экрана.

14. О безналичных расчетах в Российской Федерации: Положение Банка России № 2-П от 03.10.02 г. – Режим доступа: http://www. cbr.ru, свободный. – Загл. с экрана.

15. О порядке ведения кассовых операций в кредитных организациях на территории РФ: положение Банка России № 199-17 от 09.10.2002 г. – Режим доступа: http://www.cbr.ru, свободный. – Загл. с экрана.

16. О Центральном банке Российской Федерации (Банке России): Федеральный закон №86-ФЗ от 10.07.02 г. – Режим доступа: http://www.cbr.ru, свободный. – Загл. с экрана.

17. О банках и банковской деятельности: Федеральный закон № 17-ФЗ от 03.02.96 (с изменениями и дополнениями). – Режим доступа: http://www.cbr.ru, свободный. – Загл. с экрана.

18. О рынке ценных бумаг [текст]: официальный текст. Федеральный закон РФ от 22. 04. 1996г. №160-ФЗ.//Информационно-правовая система «Консультант Плюс».

19. Указание Банка России от 18.01.2016 №3939-У «О неприменении отдельных нормативных актов Федеральной комиссии по рынку ценных бумаг». – Режим доступа: http://www.cbr.ru, свободный. – Загл. с экрана.

20. Указание Банка России от 24.11.2016 №4212-У «О перечне, формах и порядке составления и представления форм отчетности кредитных организаций в Центральный Банк Российской Федерации // Справочно-правовая система «КонсультантПлюс».

21. Федеральный закон от 27 июня 2011 года № 161-ФЗ «О национальной платежной системе» Положением Банка России от 2 мая 2012 года № 378-П «О порядке направления в Банк России заявления о регистрации оператора платежной системы», Реестр операторов платежных систем: https://cbr.ru/registries/nps/rops/

22. Указание Банка России «О внесении изменений в Положение Банка России от 30 мая 2014 года № 421-П «О порядке расчета показателя краткосрочной ликвидности («Базель III»)»: http://www.cbr.ru/queries/unidbquery/file/90134?fileid=895

23. Положением Банка 2 России от 11 января 2021 года № 753-П «Об обязательных резервах кредитных организаций»: https://cbr.ru/Queries/UniDbQuery/File/90134/2756

24. № 4927-У от 08.10.2018 Указание Банка России «О перечне, формах и порядке составления и представления форм отчетности кредитных организаций в Центральный банк Российской Федерации» – Режим доступа: http://www.cbr.ru, свободный. – Загл. с экрана.

25. Об электронной подписи : Федеральный закон № 63-фз от 6 апреля 2011 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

26. О персональных данных : Федеральный закон № 152-фз от 27 июля 2006 года [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

27. Об инфраструктуре, обеспечивающей информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных услуг в электронной форме: постановление Правительства РФ от 8 июня 2011 г. № 451 [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

28. О федеральных государственных информационных системах, обеспечивающих предоставление в электронной форме государственных и муниципальных услуг (осуществление функций) : постановление Правительства РФ от 24 октября 2011 г. № 861 [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

29. О мерах по обеспечению перехода федеральных органов исполнительной власти и органов государственных внебюджетных фондов на межведомственное информационное взаимодействие в электронном виде: постановление Правительства РФ от 28 декабря 2011 г. № 1184 [Электронный ресурс] Доступ из справ.-правовой системы «КонсультантПлюс».

30. Указ Президента Российской Федерации от 31.03.2023 № 231 «О создании, развитии и эксплуатации государственных информационных систем с использованием единой цифровой платформы Российской Федерации «ГосТех» [Электронный ресурс] https://platform.gov.ru/documents/

31. Приказ Федерального казначейства от 12 мая 2017 г. N 11н "Об утверждении Порядка ведения Государственной информационной системы о государственных и муниципальных платежах" [Электронный ресурс] Доступ из справ.-равовой системы «Гарант» https://base.garant.ru/71729516/

7.4 Интернет-ресурсы, справочные системы

1. http://www.minfin.ru – Официальный сайт Министерства финансов РФ.
2. http://www.roscazna.ru – Официальный сайт Казначейства России.
3. http://www.eLIBRARY.ru – научная электронная библиотека.

4. http://www.budget.gov.ru – Единый портал бюджетной системы РФ (Электронный бюджет).
5. http://www.datamarts.roskazna.ru – Информационно-аналитическая система ключевых показателей исполнения бюджета.
6. http://e.lanbook.com - ЭБС «Лань»
7. https://znanium.ru/site/request-password-reset - ЭБС «Znanium»
8. http://www.iprbookshop.ru - ЭБС «IPRBooks»
9. http://www.nalog.ru - Сайт Федеральной налоговой службы РФ
10. Всероссийский НИИ проблем вычислительной техники и информатизации Министерства Российской Федерации по связи и информатизации (ВНИИПВТИ). Режим доступа: http://www.vniipvti.ru.
11. Концепция государственной информационной политики. Режим доступа: http://www.duma.gov.ru.
12. Методические рекомендации «Инфраструктурные технологические сервисы единой цифровой платформы Российской Федерации «ГосТех». Основные требования к составу и функциям»: утверждено протоколом Президиума Правительственной комиссии от 10.01.2024 № 1пр [Электронный ресурс] https://platform.gov.ru/documents/
13. Методические рекомендации о регламенте (стандарте) управления данными государственных информационных систем, создаваемых, развиваемых и эксплуатируемых на единой цифровой платформе Российской Федерации «ГосТех»: утверждено протоколом Президиума Правительственной комиссии от 15.12.2023 № 58пр [Электронный ресурс] https://platform.gov.ru/documents/
14. Методические рекомендации по проектированию и оптимизации бизнес-процессов государственных функций, услуг и сервисов, подлежащих реализации на Единой цифровой платформе Российской Федерации «ГосТех»: Утверждено протоколом Президиума Правительственной комиссии от 20.10.2023 № 47пр [Электронный ресурс] https://platform.gov.ru/documents/

7.5 Иные источники

1. Камолов, С. Г. Цифровое государственное управление : учебник для вузов / С. Г. Камолов. — Москва : Издательство Юрайт, 2023. — 336 с. — (Высшее образование). — ISBN 978-5-534-14992-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/520044 (вход по паролю).
2. Информационные технологии в экономике и управлении в 2 ч. Часть 1 : учебник для вузов / В. В. Трофимов [и др.] ; под редакцией В. В. Трофимова. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 269 с. — (Высшее образование). — ISBN 978-5-534-09083-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/517142 (вход по паролю).
3. Информационные технологии в экономике и управлении в 2 ч. Часть 2 : учебник для вузов / В. В. Трофимов [и др.] ; под редакцией В. В. Трофимова. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 245 с. — (Высшее образование). — ISBN 978-5-534-09084-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/517144 (вход по паролю).
4. Сбитнева, Г. И. Отраслевые информационные ресурсы. Практикум : учебное пособие для вузов / Г. И. Сбитнева. — 2-е изд. — Москва : Издательство Юрайт, 2022. — 154 с. — (Высшее образование). — ISBN 978-5-534-14441-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/496996 (вход по паролю).
5. Пынько, Л. Е. Применение нейросетевого анализа в оценке регионального управления цифровизацией экономики / Л. Е. Пынько, Е. В. Толкачева. – Хабаровск : Дальневосточный институт управления - филиал Федерального государственного бюджетного образовательного учреждения высшего профессионального образования

"Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации", 2020. – 87 с. – ISBN 978-5-94456-285-2. – EDN BQYOAG.

Базы данных, информационно-справочные и поисковые системы

1. www.biblio-online.ru –Электронно-библиотечная система [ЭБС] Юрайт.
2. http://www.iprbookshop.ru – Электронно-библиотечная система [ЭБС] «Iprbooks»
3. https://e.lanbook.com - Электронно-библиотечная система [ЭБС] «Лань».
4. http://elibrary.ru/ - Научная электронная библиотекаElibrary.ru.
5. https://new.znanium.com Электронно-библиотечная система [ЭБС] «Znanium.com».

6 Справочно-правовая система «Консультант».

7. Электронный периодический справочник «Гарант».

8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Адрес	Наименование оборудованных учебных кабинетов	Оснащенность оборудованных учебных кабинетов, перечень программного обеспечения
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Учебные аудитории для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ), практик, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.

680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Специализированные аудитории, оборудованные для проведения занятий по информационным технологиям (компьютерные классы)	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доска маркерная, учебные пособия, тренажеры и другие средства и оборудование. Аудитории обеспечены компьютерной техникой с возможностью подключения к сети «Интернет» и электронной информационно-образовательной среде организации. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение, справочно-правовая система «Консультант плюс», информационно-справочная система «Кодекс», Any Logic University, KNIME (Open Source программное обеспечение с открытым исходным кодом), Gretl, Statgraphics Centurion XV Academic Pro, 1С Предприятие 8 для учебных заведений, Eviews Academic, Adobe Photoshop
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Учебные аудитории для самостоятельной работы обучающихся	Специализированная мебель для организации рабочих мест обучающихся, технические средства обучения, мультимедийное оборудование, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Специализированная аудитория, оборудованная для проведения занятий по криминалистике / кабинет криминалистики	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доска маркерная, макеты, реактивы, раздаточные методические материалы и наглядные пособия, реквизит к осмотру места происшествия, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.

680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33, Точка кипения	Центр (класс) деловых игр	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, учебно-наглядные пособия, доска маркерная, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Специальная библиотека (библиотека ограниченного доступа) /	Помещение для самостоятельной работы обучающихся. Специализированная мебель для организации рабочих мест обучающихся, технические средства обучения, мультимедийное оборудование. Читальный зал научной библиотеки оснащен компьютерной техникой с возможностью подключения к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Кабинет иностранных языков	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доска маркерная, макеты, наглядные учебные пособия, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Научно-исследовательская лаборатория	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доска маркерная, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.

680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Помещения для научно-исследовательской работы	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доска маркерная, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.
680000, Хабаровский край, г. Хабаровск, ул. Муравьева-Амурского, д.33	Актовый зал	Специализированная мебель для организации рабочих мест обучающихся и преподавателя, технические средства обучения, мультимедийное оборудование, доступ к сети «Интернет» и электронной информационно-образовательной среде. Программное обеспечение: Microsoft Windows, пакет Microsoft office, антивирусное программное обеспечение.

Информационные технологии реализуются через электронно-информационную образовательную среду академии.