

Федеральное государственное образовательное бюджетное
учреждение высшего образования
«Финансовый университет при Правительстве Российской
Федерации»
(Финансовый университет)
Колледж информатики и программирования

УТВЕРЖДАЮ

Заместитель директора по
учебной работе

 Н.Ю. Долгова
« 26 » июле 2023г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ОП.14 ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ

09.02.07 Информационные системы и программирование

Москва 2023г.

Рабочая программа дисциплины разработана на основе федерального государственного образовательного стандарта среднего профессионального образования (далее – ФГОС СПО) по специальности 09.02.07 Информационные системы и программирование.

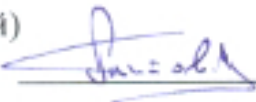
Разработчики:

Горланов Владимир Владимирович, преподаватель Колледжа информатики и программирования

Рабочая программа учебной дисциплины рассмотрена и рекомендована к утверждению на заседании предметной (цикловой) комиссии информационных систем и программирования
(наименование ПЦК)

Протокол от « 11 » мая 2023г. № 2

Председатель предметной (цикловой)
комиссии

 Н.Г. Титов

1. Общая характеристика рабочей программы учебной дисциплины

1.1. Место дисциплины в структуре основной образовательной программы

Учебная дисциплина «ОП.14 Основы машинного обучения» является вариативной частью общепрофессионального цикла основной профессиональной образовательной программы в соответствии с ФГОС по специальности 09.02.07 Информационные системы и программирование.

Учебная дисциплина «ОП.14 Основы машинного обучения» обеспечивает формирование профессиональных и общих компетенций по всем видам деятельности ФГОС специальности 09.02.07 Информационные системы и программирование.

1.2. Цель и планируемые результаты освоения дисциплины:

В рамках программы учебной дисциплины студентами осваиваются умения и знания

Код общих и профессиональных компетенций	Умения	Знания
ОК 01 ОК 02 ОК 03 ОК 04 ОК 09 ОК 10 ПК 6.5 ПК 10.1 ПК 10.2 ПК 11.1	– работать с различными источниками данных: CSV, XML и XLS; – подготавливать данные для анализа; – визуализировать результаты анализа; – выбирать оптимальный алгоритм для анализа; – использовать язык R для решения задач машинного обучения; – применять на практике алгоритмы машинного обучения для решения аналитических задач; – создавать аналитические панели; – работать с нейронными сетями.	– языка Python для анализа данных и машинного обучения; – библиотеки NumPy; – библиотеки Pandas; – библиотеки Matplotlib; – среды программирования Jupyter; – основные концепции анализа данных и машинного обучения; – основ языка программирования R; – алгоритмов и задач машинного обучения; – нейронных сетей.

2. Структура и содержание учебной дисциплины

2.1. Объем дисциплины и виды учебной работы

Вид учебной работы	Объем в часах
Объем образовательной программы учебной дисциплины	112
Объем работы студентов во взаимодействии с преподавателем	94
в том числе:	
теоретическое обучение	52
практические занятия	42
лабораторные занятия	-
контрольные работы	-
курсовая работа (проект) <i>(если предусмотрено)</i>	-
самостоятельная работа	6
консультации	-
Промежуточная аттестация в форме экзамена	12

2.2. Тематический план и содержание учебной дисциплины

Наименование разделов и тем	Содержание учебного материала и формы организации деятельности студентов	Объем в часах	Коды компетенций, формированию которых способствует элемент программы
1	2	3	4
Раздел 1. Начало анализа данных		30	
Тема 1.1. Основы анализа данных	Содержание учебного материала	12	ОК 01-03, ОК 09 ПК 10.1, ПК 11.1,
	1. Основные концепции анализа данных. 2. Основы работы с Jupyter Notebook. 3. Библиотека NumPy. Полезные инструменты. 4. Библиотека Pandas. Возможности для Data Science.	4	
	В том числе практических занятий:	8	
	1. Практическое занятие «Использование библиотеки NumPy».	4	
	2. Практическое занятие «Использование библиотеки Pandas».	4	
Тема 1.2. Предобработка данных	Содержание учебного материала	6	ОК 01-03, ОК 09, ОК 10, ПК 10.1, ПК 11.1
	1. Очистка данных от выбросов, пропусков и дубликатов 2. Преобразование разных форматов данных	4	
	В том числе практических занятий:	2	
	3. Практическое занятие «Анализ клиентов банка».	2	
Тема 1.3. Исследовательский и статистический анализ данных	Содержание учебного материала	12	ОК 01-03, ОК 09 ПК 10.1, ПК 11.1,
	1. Исследование основных свойств данных, поиск закономерностей, распределений и аномалий 2. Библиотеки SciPy и Matplotlib 3. Анализ взаимосвязей в данных методами статистики	6	
	В том числе практических занятий:	6	
	4. Практическое занятие «Анализ популярности заправок».	4	
	5. Практическое занятие «Оптимизация воронок продаж для ускорения работы отдела маркетинга».	2	
Раздел 2.	Основы машинного обучения	46	

Тема 2.1. Введение в машинное обучение	Содержание учебного материала	12	ОК 01-03, ОК 09 ПК 10.1, ПК 11.1,
	1. Основные концепции машинного обучения	4	
	2. Задачи классификации и регрессии		
	В том числе практических занятий:	8	
	6. Практическое занятие «Создание первого проекта с машинным обучением»	4	
Тема 2.2. Вспомогательные инструменты Data Science	7. Практическое занятие «Прогноз вероятности ухода клиента из банка»	4	ОК 01-03, ОК 09, ПК 6.5, ПК 10.1, ПК 10.2, ПК 11.1,
	Содержание учебного материала	4	
Тема 2.3. Математика машинного обучения	1. Работа с bash, virtualenv, docker. 2. Управление git-репозиторием	4	ОК 01-02, ОК 09
	Содержание учебного материала	12	
	1. Алгоритмы и структуры данных: сложность алгоритма, алгоритмы на графах, динамическое программирование	8	
	2. Линейная алгебра: векторы, матрицы, расстояния		
	3. Численные методы: приближенные алгоритмы, алгоритмы оптимизации, градиентный спуск		
Тема 2.4. Системы хранения данных	4. Алгоритмы машинного обучения: решающие деревья, бустинг и бэггинг, линейные и модели		ОК 01-04, ОК 09, ОК 10, ПК 6.5, ПК 10.1, ПК 10.2, ПК 11.1,
	В том числе практических занятий:	4	
	8. Практическое занятие «Метод преобразования данных для защиты личной информации клиентов»	2	
	9. Практическое занятие «Разработка модели для определения стоимости автомобиля с пробегом»	2	
	Содержание учебного материала	6	
	1. Анализ данных на SQL 2. Методы библиотеки PySpark	4	
	В том числе практических занятий:	2	
	10. Практическое занятие «Анализ спроса на авиабилеты в города, где проводятся фестивали»	2	
	Содержание учебного материала	4	ОК 01-03, ОК 09

Тема 2.5. Обучение без учителя	1. Задачи кластеризации 2. Поиск аномалий	4	ПК 10.1
Тема 2.6. Машинное обучение для текстов	Содержание учебного материала	8	ОК 01-03, ОК 09 ПК 10.1, ПК 10.2, ПК 11.1,
	1. Алгоритм TF-IDF 2. Языковые представления word2vec и BERT	2	
	В том числе практических занятий:	6	
	11. Практическая работа «Оценка токсичности комментариев»	6	
Раздел 3. Основы deep learning		18	
Тема 3.1. Компьютерное зрение	Содержание учебного материала	22	ОК 01-04, ОК 09, ОК 10, ПК 10.1, ПК 11.1,
	1. Нейронные сети 2. Метод градиентного спуска 3. Регуляризация нейронных сетей 4. Свёрточные нейронные сети 5. Библиотека Keras 6. Библиотека Tensorflow	12	
	В том числе практических занятий:	6	
	12. Практическое занятие «Построение модели анализа фотографии»	2	
	13. Практическое занятие «Оптимизация нейронной сети»	4	
Консультация		-	
Промежуточная аттестация в форме экзамена		12	
Всего:		112	

3. Условия реализации программы дисциплины

3.1. Материально-техническое обеспечение

Для реализации программы учебной дисциплины должны быть предусмотрены следующие специальные помещения (в соответствии с ФГОС и ПООП): лаборатория программирования и баз данных

- оборудование учебного кабинета: 25 рабочих мест кабинета, рабочее место преподавателя (АРМ), парты для обучающихся.
- технические средства обучения: мультимедиа-проектор, ПК с подключением к сети Интернет.

- оборудование лаборатории и рабочих мест лаборатории: ПК по количеству учащихся, подключение к локальной сети, подключение к сети Интернет.

Лицензионное программное обеспечение общего и профессионального назначения.

3.2. Информационное обеспечение реализации программы

Для реализации программы библиотечный фонд образовательной организации должен иметь печатные и/или электронные образовательные и информационные ресурсы, рекомендуемых для использования в образовательном процессе

3.2.1. Печатные издания

Законодательные и нормативные акты:

1. ГОСТ 7.1. – 2003. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. – М.: ИПК Издательство стандартов, 2004. – 169 с.
2. ГОСТ 7.32 – 2001. Система стандартов по информацию, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления. – М.: ИПК Издательство стандартов, 2001. – 21 с.
3. ГОСТ 7.82-2001 Библиографическая запись. Библиографическое описание электронных ресурсов. Общие требования и правила составления. – М.: ИПК Издательство стандартов, 2001. – 21 с.
4. Единая система программной документации. – М.: Стандартинформ, 2005. – 128 с.

3.2.2. Основная литература:

1. Кондрашов, Ю.Н., Анализ данных и машинное обучение на платформе MS SQL Server : учебное пособие / Ю.Н. Кондрашов. — Москва: Русайнс, 2021. — 303 с. — ISBN 978-5-4365-7924-5. —

URL:<https://book.ru/book/941049> (дата обращения: 08.06.2022). — Текст : электронный.

2. Воронина, В. В. Теория и практика машинного обучения: учебное пособие / В. В. Воронина. — Ульяновск: УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053> (дата обращения: 08.06.2022). — Режим доступа: для авториз. пользователей.

3.2.3. Электронные издания (электронные ресурсы):

1. Распределенные представления слов и фраз Миколов Томас. — [Электронный ресурс]. — Режим доступа: <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality>
2. Draw.io — [Электронный ресурс]. — Режим доступа: <https://startpack.ru/application/draw-io663>
3. Инструкция по работе с TensorFlow Object Detection API. — [Электронный ресурс]. — Режим доступа: <https://habr.com/company/nixsolutions/blog>
4. Многопоточность на примерах – модуль threading. — [Электронный ресурс]. — Режим доступа: <https://python-scripts.com/threading>
5. Метаклассы и метапрограммирование в Python. — [Электронный ресурс]. — Режим доступа: <https://gitjournal.tech/metaklassy-i-metaprogrammirovanie-v-python/>
6. Keras: the Python deep learning API. — [Электронный ресурс]. — Режим доступа: <https://keras.io/>
7. Библиотеки для глубокого обучения: Keras. — [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/company/ods/blog/325432/>
8. Методы оптимизации нейронных сетей. — [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/post/318970/>

4. Контроль и оценка результатов освоения учебной дисциплины

Контроль и оценка результатов освоения учебной дисциплины осуществляется преподавателем в процессе проведения практических занятий и лабораторных работ, тестирования, а также выполнения обучающимися индивидуальных заданий, проектов, исследований.

Результаты обучения	Критерии оценки	Методы оценки
<i>Перечень умений, осваиваемых в рамках дисциплины:</i> – работать с различными источниками данных: CSV, XML и XLS; – подготавливать данные для анализа; – визуализировать результаты анализа; – выбирать оптимальный алгоритм для анализа; – использовать язык R для решения задач машинного обучения; – применять на практике алгоритмы машинного обучения для решения аналитических задач; – создавать аналитические панели; – работать с нейронными сетями.	«Отлично» - теоретическое содержание курса освоено полностью, без пробелов, умения сформированы, все предусмотренные программой учебные задания выполнены, качество их выполнения оценено высоко. «Хорошо» - теоретическое содержание курса освоено полностью, без пробелов, некоторые умения сформированы недостаточно, все предусмотренные программой учебные задания выполнены, некоторые виды заданий выполнены с ошибками. «Удовлетворительно» - теоретическое содержание курса	– компьютерное тестирование по отдельным темам дисциплины; – текущий контроль в форме защиты практических работ; – экзамен по дисциплине.

<i>Перечень знаний, осваиваемых в рамках дисциплины:</i> – язык Python для анализа данных и машинного обучения; – библиотека NumPy; – библиотека Pandas; – библиотека Matplotlib; – среда программирования Jupyter; – основные концепции анализа данных и машинного обучения; – основ языка программирования R; – алгоритмы и задачи машинного обучения; – нейронные сети.	освоено частично, но пробелы не носят существенного характера, необходимые умения работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий содержат ошибки. «Неудовлетворительно» - теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки.	
---	--	--

1. Общая характеристика рабочей программы дисциплины

1.1. Место дисциплины в структуре основной образовательной программы

Учебная дисциплина «ОП.14 Основы машинного обучения» является вариативной частью общепрофессионального цикла основной профессиональной образовательной программы в соответствии с ФГОС по специальности 09.02.07 Информационные системы и программирование.

1.2. Цель и планируемые результаты освоения дисциплины:

В рамках программы дисциплины студентами осваиваются умения и знания

Код общих и профессиональных компетенция	Умения	Знания
ОК 01 ОК 02 ОК 03 ОК 04 ОК 09 ПК 6.5 ПК 10.1 ПК 10.2 ПК 11.1	– работать с различными источниками данных: CSV, XML и XLS; – подготавливать данные для анализа; – визуализировать результаты анализа; – выбирать оптимальный алгоритм для анализа; – применять на практике алгоритмы машинного обучения для решения аналитических задач; – создавать аналитические панели; – работать с нейронными сетями.	– языка Python для анализа данных и машинного обучения; – библиотеки NumPy; – библиотеки Pandas; – библиотеки Matplotlib; – среды программирования Jupyter; – основные концепции анализа данных и машинного обучения; – алгоритмов и задач машинного обучения; – нейронных сетей.

2. Структура и содержание дисциплины

2.1. Объем дисциплины и виды учебной работы

Вид учебной работы	Объем в часах
Объем образовательной программы дисциплины	107
Объем работы студентов во взаимодействии с преподавателем	92
в том числе:	
теоретическое обучение	46
практические занятия	46
лабораторные занятия	-
контрольные работы	-
Курсовой проект (работа)(если предусмотрено)	-
самостоятельная работа	3
Консультации	4
Промежуточная аттестация в форме экзамена	8

2.2. Тематический план и содержание дисциплины

Наименование разделов и тем	Содержание учебного материала и формы организации деятельности студентов	Объем в часах	Коды компетенций, формированию которых способствует элемент программы
1	2	3	4
Раздел 1. Начало анализа данных		26	
Тема 1.1. Основы анализа данных	Содержание учебного материала	10	ОК 01 ОК 02 ОК 03 ОК 09 ПК 10.1 ПК 11.1
	1. Основные концепции анализа данных.	1	
	2. Основы работы с Jupyter Notebook.	1	
	3. Библиотека NumPy. Полезные инструменты.	2	
	4. Библиотека Pandas. Возможности для Data Science.	2	
	В том числе практических занятий:	4	
	1. Практическое занятие «Использование библиотеки NumPy».	2	
	2. Практическое занятие «Использование библиотеки Pandas».	2	
	Самостоятельная работа студентов	-	
Тема 1.2. Предобработка данных	Содержание учебного материала	6	ОК 01 ОК 02 ОК 03 ОК 09 ПК 10.1 ПК 11.1
	1. Очистка данных от выбросов, пропусков и дубликатов	2	
	2. Преобразование разных форматов данных	2	
	В том числе практических занятий:	2	
	1. Практическое занятие «Анализ клиентов банка».	2	
	Самостоятельная работа студентов	-	
Тема 1.3. Исследовательский и статистический анализ данных	Содержание учебного материала	10	ОК 01 ОК 02 ОК 03 ОК 09 ПК 10.1 ПК 11.1
	1. Исследование основных свойств данных, поиск закономерностей, распределений и аномалий	2	
	2. Библиотеки SciPy и Matplotlib	2	
	3. Анализ взаимосвязей в данных методами статистики	2	
	В том числе практических занятий:	4	
	1. Практическое занятие «Анализ популярности заправок».	2	
	2. Практическое занятие «Анализ воронок продаж для оптимизации работы отдела маркетинга».	2	
	Самостоятельная работа студентов	-	
Раздел 2. Основы машинного обучения		45	

	В том числе практических занятий:	2	ОК 04 ОК 09 ПК 6.5 ПК 10.1 ПК 10.2 ПК 11.1
	1. Практическое занятие «Анализ данных при помощи SQL»	2	
	Самостоятельная работа студентов	-	
Тема 2.5. Обучение без учителя	Содержание учебного материала	6	ОК 01 ОК 02 ОК 03 ОК 09 ПК 10.1 ПК 11.1
	1. Задачи кластеризации Поиск аномалий	2	
	В том числе практических занятий:	4	
	1. Практическая работа «Сегментация клиентов»	2	
	2. Практическая работа «Поиск аномалий методами машинного обучения»	2	
	Самостоятельная работа студентов	-	
Тема 2.6. Машинное обучение для текстов	Содержание учебного материала	8	ОК 01 ОК 02 ОК 03 ОК 09 ПК 10.1 ПК 11.1
	Алгоритм TF-IDF и языковое представление word2vec	2	
	В том числе практических занятий:	6	
	1. Практическая работа «Обнаружение токсичных комментариев»	2	
	2. Практическая работа «Обнаружение спама»	2	
	3. Практическая работа «Классификация новостей по категориям»	2	
	Самостоятельная работа студентов	-	
Раздел 3. Основы Deep Learning		24	
Тема 3.1. Компьютерное зрение	Содержание учебного материала	24	ОК 01 ОК 02 ОК 03 ОК 09 ПК 10.1 ПК 11.1
	1. Нейронные сети	2	
	2. Метод градиентного спуска	2	
	3. Регуляризация нейронных сетей	2	
	4. Свёрточные нейронные сети	2	
	5. Библиотека Keras	2	
	6. Библиотека TensorFlow	2	
	В том числе практических занятий:	12	
	1. Практическое занятие «Построение модели нейронной сети»	6	

	2. Практическое занятие «Оптимизация нейронной сети»	6	
	Самостоятельная работа студентов	-	
Консультация		4	
Промежуточная аттестация в форме экзамена		8	
Всего:		107	

3. Условия реализации дисциплины

3.1. Для реализации программы дисциплины должны быть предусмотрены следующие специальные помещения: лаборатория программирования и баз данных

- оборудование учебного кабинета: 25 рабочих мест кабинета, рабочее место преподавателя (АРМ), парты для обучающихся.
- технические средства обучения: мультимедиа-проектор, ПК с подключением к сети Интернет.
- оборудование лаборатории и рабочих мест лаборатории: ПК по количеству учащихся, подключение к локальной сети, подключение к сети Интернет.
- Лицензионное программное обеспечение общего и профессионального назначения.

3.2. Информационное обеспечение реализации программы

Для реализации программы библиотечный фонд структурного подразделения должен иметь печатные и/или электронные образовательные и информационные ресурсы для использования в образовательном процессе. При формировании библиотечного фонда выбирается не менее одного издания из перечисленных в ПООП печатных изданий и (или) электронных изданий в качестве основного, при этом список, может быть дополнен новыми изданиями.

Основные печатные и электронные издания:

1. Кондрашов, Ю.Н., Анализ данных и машинное обучение на платформе MS SQL Server : учебное пособие / Ю.Н. Кондрашов. — Москва: Русайнс, 2024. — 303 с. — ISBN 978-5-466-06473-5. — URL: <https://book.ru/book/941049>
2. Коротеев, М.В., Основы машинного обучения на Python: учебник / М.В. Коротеев. — Москва: КиноРус, 2024. — 431 с. — ISBN 978-5-406-12673-8. — URL: <https://book.ru/book/952751>

Дополнительные источники:

1. Распределенные представления слов и фраз Миколов Томас. — [Электронный ресурс]. — Режим доступа: <https://papers.nips.cc/paper/5021>
2. Draw.io — [Электронный ресурс]. — Режим доступа: <https://startpack.ru/application/draw-io663>
3. Инструкция по работе с TensorFlow Object Detection API. — [Электронный ресурс]. — Режим доступа: <https://habr.com/company/nixsolutions/blog>
4. Многопоточность на примерах — модуль threading. — [Электронный ресурс]. — Режим доступа: <https://python-scripts.com/threading>

5. Метаклассы и метапрограммирование в Python. – [Электронный ресурс]. – Режим доступа: <https://gitjournal.tech/metaklassy-imetaprogrammirovanie-v-python/>
6. Keras: the Python deep learning API. – [Электронный ресурс]. – Режим доступа: <https://keras.io/>
7. Библиотеки для глубокого обучения: Keras. – [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/company/ods/blog/325432/>
8. Методы оптимизации нейронных сетей. – [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/318970/>

4. Контроль и оценка результатов освоения дисциплины

Контроль и оценка результатов освоения дисциплины осуществляется преподавателем в процессе проведения практических занятий, тестирования, а также выполнения студентами индивидуальных заданий, проектов, исследований.

Результаты обучения	Критерии оценки	Методы оценки
Перечень знаний, осваиваемых в рамках дисциплины: – языка Python для анализа данных и машинного обучения; – библиотеки NumPy; – библиотеки Pandas; – библиотеки Matplotlib; – среды программирования Jupyter; – основные концепции анализа данных и машинного обучения; – алгоритмов и задач машинного обучения; – нейронных сетей; – работать с различными источниками данных: CSV, XML и XLS; – подготавливать данные для анализа; – визуализировать результаты анализа; – выбирать оптимальный алгоритм для анализа; – применять на практике алгоритмы машинного обучения для решения аналитических задач; – создавать аналитические панели; – работать с нейронными сетями.	«Отлично» - теоретическое содержание курса освоено полностью, без пробелов, умения сформированы, все предусмотренные программой учебные задания выполнены, качество их выполнения оценено высоко. «Хорошо» - теоретическое содержание курса освоено полностью, без пробелов, некоторые умения сформированы недостаточно, все предусмотренные программой учебные задания выполнены, некоторые виды заданий выполнены с ошибками. «Удовлетворительно» - теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые умения работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий содержат ошибки. «Неудовлетворительно» - теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки.	– компьютерное тестирование по отдельным темам дисциплины; – текущий контроль в форме защиты практических работ; – экзамен по дисциплине.