


Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Кубанский государственный университет»
Экономический факультет

УТВЕРЖДАЮ:
Проректор по учебной работе,
качеству образования,
первый проректор
Хагуров Т.А.
« 29 » 2020 г.



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.09 ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ**

Направление	38.04.08 Финансы и кредит
Профиль	Финансы в цифровой экономике
Программа подготовки	Академическая
Форма обучения	Очная
Квалификация (степень) выпускника	Магистр

Краснодар 2020

1 Цели и задачи изучения дисциплины

1.1 Цель дисциплины

Целью дисциплины является ознакомление студентов с основополагающими элементами современного анализа больших наборов данных. Самым перспективным подходом к анализу данных считается применение машинного обучения (Machine Learning) — набора методов, благодаря которым компьютер может находить в массивах изначально неизвестные взаимосвязи и закономерности.

1.2 Задачи дисциплины

- Дать теоретические знания в области машинного обучения и искусственного интеллекта;
- Изучить математические основы машинного обучения и источники информации для проведения финансово-экономических расчетов;
- Освоить базовые методы машинного обучения;
- Научить строить собственные модели машинного обучения, выдвигать концепции, подтверждать и опровергать их;
- Показать основные проблемы, возникающие при моделировании и способы их преодоления;
- Развивать способности студентов анализировать и объективно интерпретировать результаты моделирования;
- Научить эффективно использовать инструменты Data Science в финансах.
- Привить студентам навыки получать, хранить и перерабатывать большие данные, применять полученные в ходе изучения дисциплины знания при проведении научно-исследовательской работы, подготовки магистерской диссертации.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина "Основы машинного обучения" относится к обязательным в вариативной части Блока 1 учебного плана по направлению подготовки 38.04.08 «Финансы и кредит» (квалификация (степень) «магистр»). Дисциплина предназначена для студентов второго года обучения магистерской программы «Финансы в цифровой экономике», читается в 4-м семестре 2-го курса.

Дисциплина опирается на общие математические и экономические знания, умения и навыки, получаемые в процессе освоения учебных курсов: "Высшая математика", "Теория вероятностей и математическая статистика", "Информатика", "Статистика", "Эконометрика". Студент должен уметь работать с математической и экономической литературой, владеть основными навыками получения, хранения, переработки информации, иметь хорошие навыки работы с компьютером, в том числе с глобальными информационными сетями, а также программами для обработки статистических данных, позволяющими производить первичный анализ данных (Statistica, EViews и т.п.). Кроме того, студент должен иметь представление о программировании, вла-

деть базовыми навыками написания программ на одном из высокоуровневых языков, приспособленных для работы с векторами и массивами, типа octave, python.

Знания, умения и навыки, полученные студентами в процессе изучения дисциплины, используются для прохождения преддипломной практики, а также при написании магистерской диссертации.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся профессиональных компетенций.

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			Знать	Уметь	Владеть
1.	ПК-2	Способность анализировать и использовать различные источники информации для проведения финансово-экономических расчетов	Общепринятую терминологию, типовые задачи машинного обучения и существующие методы их решения.	Использовать современные вычислительные средства инструментальные средства и среды для анализа данных различных источников информации для проведения финансово-экономических расчетов.	Основными алгоритмами машинного обучения для анализа различные источники информации для проведения финансово-экономических расчетов.
2.	ПК-18	Способность осуществлять разработку инструментов проведения исследований в области финансов и кредита, анализ их результатов, подготовку данных для составления финансовых обзоров, отчетов и научных публикаций	Основные возможности существующего инструментария, применяемого для машинного обучения и анализа данных для исследований в области финансов и кредита	Делать правильные выводы из сопоставления результатов теории и эксперимента	Навыками постановки задач классификации, кластеризации, прогнозирования для исследований в области финансов и кредита, анализ их результатов, подготовку данных для составления финансовых обзоров, отчетов и научных публикаций
3.	ПК-20	Способность осу-	Виды задач,	Разрабатывать	Приёмами

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			Знать	Уметь	Владеть
		способствовать разработке теоретических и новых эконометрических моделей исследуемых процессов, явлений и объектов, относящихся к сфере профессиональной финансовой деятельности в области финансов и кредита, давать оценку и интерпретировать полученные в ходе исследования результаты	решаемых методами машинного обучения, и выбирать подходящие методы для исследования явлений и объектов, относящихся к сфере профессиональной финансовой деятельности в области финансов и кредита, давать оценку и интерпретировать полученные в ходе исследования результаты	модели машинного обучения с использованием различных инструментов для исследования явлений и объектов в сфере финансовой деятельности в области финансов и кредита.	оценки эффективности моделей анализа данных

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 3 зач.ед. (108 часов), их распределение по видам работ представлено в таблице

Вид учебной работы	Всего часов	Семестры
		4
Контактная работа, в том числе	32,3	32,3
Аудиторные занятия (всего)	32	32
В том числе:		
Занятия лекционного типа	8	8
Занятия семинарского типа (семинары, практические занятия)	24	24
Лабораторные занятия	-	-
Иная контактная работа	0,3	0,3
Промежуточная аттестация (ИКР)	0,3	0,3
Самостоятельная работа (всего)	49	49
В том числе:		
Проработка учебного теоретического материала (Самостоятельное изучение разделов)	34	34
Подготовка к текущему контролю	15	15
Вид промежуточной аттестации (экзамен)	26,7	26,7
Общая трудоёмкость	108	108
Контактная работа	32,3	32,3
зач. ед.	3	3

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы дисциплины, изучаемые в 4 семестре

№ раздела	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Обзор задач машинного обучения и используемый инструментарий	10	2	4	-	4
2.	Машинное обучение с учителем (регрессия и классификация данных) и без учителя (кластеризация данных)	26	2	8	-	16
3.	Улучшение качества моделей	14	2	4		8
4.	Введение в нейронные сети	31	2	8	-	21
	Итого по дисциплине:	-	8	24	-	49

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов дисциплины:

2.3.1 Занятия лекционного типа

№ раздела	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
1	Обзор задач машинного обучения и используемый инструментарий	Обзор возможностей применения машинного обучения в финансах. Основы линейной алгебры. Обзор инструментов, для работы с векторами и матрицами.	К О
2	Машинное обучение с учителем (регрессия и классификация данных) и без учителя (кластеризация данных)	Линейная регрессия. Функция стоимости. Оптимизация методом градиентного спуска. Нормальное уравнение. Логистическая регрессия - метод классификации данных на отдельные результаты. Обучение без учителя и алгоритм кластеризации k-средних.	К О
3	Улучшение качества моделей	Оценка гипотезы машинного обучения. Проблемы переобучения моделей. Регуляризация данных. Изучение кривых. Анализ ошибок.	К О
4	Введение в нейронные сети	Простейшие нейронные сети. Нейроны и слои. Обратное распространение. Функция активации. Реализация нейронной сети на высокоуровневом языке программирования. Повышение скорости работы нейронной сети. Проблемы и расширения нейронных сетей.	К О

2.3.2 Занятия семинарского типа

№ раздела	Наименование раздела	Содержание раздела	Форма текущего контроля
1	2	3	4
1	Обзор задач машинного обучения и используемый инструментарий	Решение практических задач по теме «Начисление процентов»	РГЗ
2	Машинное обучение с учителем (регрессия и классификация данных) и без учителя (кластеризация данных)	Решение практических финансовых задач с помощью моделей машинного обучения.	РГЗ
3	Улучшение качества моделей	Оптимизация построенных моделей машинного обучения	РГЗ
4	Введение в нейронные сети	Моделирование нейронной сети.	РГЗ

Р-реферат, З-задачи, РГЗ-расчетно-графическое задание, Д-дискуссия, К-коллоквиум, КС-круглый стол, Т- тестирование, О-опрос, Р- реферат, Э-Эссе по одной из тем по предлагаемой в рамках занятий тематике.

2.3.3 Лабораторные занятия

Лабораторные занятия - не предусмотрены

2.3.4 Примерная тематика курсовых работ (проектов)

Курсовые работы - не предусмотрены

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	2	3

1.	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к практическим занятиям	<p>Ресурсы (в т.ч. и электронные) библиотеки КубГУ</p> <ol style="list-style-type: none"> 1. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования: Учебное пособие / Вьюгин В.В. - Москва: МЦНМО, 2014. - 304 с.: ISBN 978-5-4439-2014-6. - Текст: электронный. - URL: https://znanium.com/catalog/product/958689 2. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / С. Рашка ; пер. с англ. А.В. Логунова. - Москва: ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0. - Текст: электронный. - URL: https://znanium.com/catalog/product/1027758 3. Кадури А. А., Николенко С.И. Глубокое обучение. Погружение в мир нейронных сетей – Москва: Питер, 2018. - 402 с.: ISBN 978-5-4439-2014-6 4. Методические указания по выполнению самостоятельной работы обучающихся. Утв. на заседании Совета экономического факультета ФГБОУ ВО «КубГУ». Протокол № 8 от 29 июня 2017 г. https://www.kubsu.ru/ru/econ/metodicheskie-ukazaniya
----	---	--

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии

В процессе изучения дисциплины лекции, практические занятия, консультации являются ведущими формами обучения в рамках лекционно-семинарской образовательной технологии.

Лекции излагаются в виде презентации, затем данные материалы в электронной форме передаются студентам.

Основной целью практических занятий является разбор практических ситуаций. Дополнительной целью практических занятий является контроль усвоения пройденного материала. На практических занятиях также осуществляется проверка выполнения заданий.

При реализации различных видов учебной работы (лекций и практических занятий) используются следующие образовательные технологии: решение задач, контрольные вопросы.

Для визуализации учебного процесса на лекциях и семинарах предполагается использование мультимедийного проектора (презентации Microsoft PowerPoint).

Для инвалидов и лиц с ограниченными возможностями здоровья устанавливается особый порядок освоения указанной дисциплины, возможна организация консультаций с использованием электронной почты.

4. Оценочные средства для текущего контроля успеваемости и итоговой аттестации

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- в печатной форме с адаптированным размером шрифта,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4.1 Фонд оценочных средств для проведения текущей аттестации

В ходе учебного процесса предусматривается систематический контроль знаний. При этом по каждому разделу предусматривается определенный тип контроля, а именно, решение задач, направленное на определение степени сформированности компетенций.

Пример задач по отдельным темам дисциплины.

1. Фирме необходим кредит на сумму 60 млн. руб. сроком на три месяца. Годовая процентная ставка (простые проценты) за кредиты (без учета инфляции) - 48%. Месячные уровни инфляции за три предыдущих месяца - 3,4%, 2,5%, 5,2%. Кредит выделен с расчетом на средний уровень инфляции, рассчитанный по трем предыдущим месяцам. Определите процентную ставку

кредита, наращенную сумму и дисконт банка.

2. Какую сумму должна ежемесячно в конце месяца переводить фирма в банк, чтобы на счете через 2 года была сумма 3 млн. руб. Номинальная процентная ставка банка при ежемесячном начислении процентов — 12%.

3. Определите текущую стоимость пакета из 50 облигаций номинальной стоимостью 0,5 млн. руб. каждая сроком погашения 3 года. Доход по облигациям выплачивается в конце срока по сложной ставке процентов - 24% годовых. Банковская процентная ставка - 22% годовых.

4. Стоимость товара по форфейтингу - 400 млн. руб. Покупатель выдал продавцу 8 векселей. Срок погашения каждого из них - 3 месяца. Проценты по векселям — 32% годовых. Продавец учитывал векселя в банке по учетной ставке 40% годовых. Определите стоимость каждого векселя, сумму портфеля векселей и сумму, полученную продавцом при его продаже.

4.2 Фонд оценочных средств для проведения промежуточной аттестации (экзамен)

Перечень вопросов для подготовки к итоговой аттестации:

1. Что такое объект, целевая переменная, признак, модель, функционал ошибки и обучение?
2. Запишите формулы для линейной модели регрессии и для среднеквадратичной ошибки. Запишите среднеквадратичную ошибку в матричном виде.
3. Что такое коэффициент детерминации? Как интерпретировать его значения?
4. Что такое градиент? Какое его свойство используется при минимизации функций?
5. Запишите формулу для одного шага градиентного спуска. Какие способы оценивания градиента вы знаете? Почему не всегда можно использовать полный градиентный спуск?
6. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?
7. Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?
8. Что такое регуляризация? Запишите L1- и L2-регуляризаторы. Почему L1-регуляризация отбирает признаки?
9. Запишите формулу для линейной модели классификации. Что такое отступ? Как обучаются линейные классификаторы и для чего нужны верхние оценки пороговой функции потерь?
10. Что такое точность, полнота и F-мера?
11. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
12. Запишите задачу метода опорных векторов для линейно неразделимого случая. Как функционал этой задачи связан с отступом классификатора?

13. В чём заключается подход с независимой классификацией в задаче классификации с пересекающимися классами?
14. В чём заключается преобразование категориальных признаков в вещественные с помощью счётчиков? Почему использование счётчиков может привести к переобучению? Какие методы борьбы с этой проблемой счётчиков вам известны?
15. Как определить для линейной модели, какие признаки являются самыми важными?
16. Задача кластеризации. Метрики качества.
17. Метод K-Means, вывод его шагов.
18. Понижение размерности. Метод главных компонент: постановка задачи и решение.
19. Постановка задачи построения рекомендаций. Коллаборативные методы: memory-based и на основе моделей со скрытыми переменными. Контентные методы.
20. Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации.
21. Проблема полноты. Задача исключаящего или полнота двухслойных сетей в пространстве булевых функций.
22. Расскажите Алгоритм обратного распространения ошибок.
23. Подбор структуры сети: методы постепенного усложнения сети.

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

5.1 Основная литература:

1. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования: Учебное пособие / Вьюгин В.В. - Москва :МЦНМО, 2014. - 304 с.: ISBN 978-5-4439-2014-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/958689>
2. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / С. Рашка ; пер. с англ. А.В. Логунова. - Москва : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027758>
3. Кадурин А. А., Николенко С.И. Глубокое обучение. Погружение в мир нейронных сетей – Москва :Питер, 2018. - 402 с.: ISBN 978-5-4439-2014-6

5.2 Дополнительная литература:

1. Миркин Б.Г. Введение в анализ данных: учебник и практикум / Б.Г. Миркин. – М.: Издательство Юрайт, 2017. – 174 с. Режим доступа: www.biblioonline.ru/book/B6C7142B-D75C-4706-B7D6-9193D1B15678
2. Ковалев, Валерий Викторович. Финансовый анализ [Текст] : управление капиталом. Выбор инвестиций. Анализ отчетности. - 2-е изд., перераб.

и доп. - М. : Финансы и статистика, 2000. - 511 с. - Библиогр.: с. 502 - 507. - ISBN 5279020435 : 70.00. 5. Кочетыгов А.А. «Финансовая математика», М., «Феникс», 2004 г.

3. Кочович, Елена. Финансовая математика с задачами и решениями [Текст] : учебно-методическое пособие / Е. Кочович ; [пер. с серб. Е. Кочович]. - Изд. 2-е, перераб. и доп. - М. : Финансы и статистика, 2004. - 380 с. - Библиогр.: с. 380. - ISBN 5279029084 : 268.80. 7. Мелкумов Я.С., Румянцев В.Н. Финансовые вычисления в коммерческих расчетах. — М.:, 1994.

4. Румянцева, Е. Е. Инвестиции и бизнес-проекты [Электронный ресурс] : учебно-практическое пособие : в 2 ч. Ч. 2 / Е. Е. Румянцева. - 2-е изд., стер. - Москва ; Берлин : Директ-Медиа, 2016. - 348 с., ил. - <http://biblioclub.ru/index.php?page=book&id=450110>.

6. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

— <https://www.gnu.org/software/octave/>

7. Методические указания для обучающихся по освоению дисциплины (модуля)

Основной формой самостоятельной работы студента по изучаемому курсу является выполнение расчетно-графического задания (РГЗ) с целью закрепление полученных студентами на лекциях теоретических знаний, моделирования практических ситуаций, а также проверка эффективности самостоятельной работы студента.

В ходе самостоятельной работы студенту необходимо изучить, прежде всего, ту учебную литературу, монографии и прочие источники информации, которые рекомендованы преподавателем.

Выполнение расчетно-графического задания предполагает следование четкому алгоритму расчетов, предлагаемых к каждому заданию и выполнению выводов по проделанным расчетам.

При решении задач следует производить расчеты с указанием соответствующих формул, приведенных в РГЗ, а также сформировать итоговые таблицы, представляющие полученные результаты.

При необходимости в процессе работы над заданием студент может получить индивидуальную консультацию у преподавателя.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

8.1 Перечень необходимого программного обеспечения

Для успешного освоения дисциплины, студент использует такие программные средства как Octave, Python.

8.2 Перечень необходимых информационных справочных систем

Не используются.

9. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

№	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащенность
1.	Занятия лекционного типа: лекции-презентации Классическая лекция	Аудитории, оснащенные презентационной техникой (проектор, экран, ноутбук) и соответствующим программным обеспечением (MicrosoftOffice). Ауд. 520А, 207Н, 208Н, 209Н, 205А, 4033Л, 4038Л, 4039Л, 5040Л, 5041Л, 5042Л, 5045Л, 5046Л Ауд. 212Н, 214Н, 201А
2.	Занятия семинарского типа	Аудитории А208Н, 202А, 210Н, 216Н, 513А, 514А, 515А, 516А, а также аудитории, оснащенные презентационной техникой (проектор, экран, ноутбук) и соответствующим программным обеспечением (MicrosoftOffice 2016). Ауд., 2026Л, 2027Л, 4034Л, 4035Л, 4036Л, 5043Л, 201Н, 202Н, 203Н, А203Н
3.	Групповые и индивидуальные консультации	Кафедра экономического анализа, статистики и финансов (ауд. 223), А208Н
4.	Текущий контроль самостоятельной работы	Аудитории, оснащенные презентационной техникой (проектор, экран, ноутбук) и соответствующим программным обеспечением (MicrosoftOffice). Ауд. 520А, 207Н, 208Н, 209Н, ,2026Л, 2027Л, 4033Л, 4034Л, 4035Л, 4036Л, 4038Л, 4039Л, 5040Л, 5041Л, 5042Л, 5043Л, 5045Л, 5046Л, 201Н, 202Н, 203Н, А203Н, 205А,
5.	Промежуточная аттестация	Ауд. 212Н, 214Н, 201А, А208Н, 202А, 210Н, 216Н, 513А, 514А, 515А, 516А
6.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета Ауд. 213А, 218А, 201Н, 202Н, 203Н, А203Н