

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ  
М.В.ЛОМОНОСОВА»**

**ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ**

**«УТВЕРЖДАЮ»**

Декан экономического факультета МГУ

профессор \_\_\_\_\_ А.А.Аузан

«\_\_\_» \_\_\_\_\_ 2020 год

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Наименование дисциплины:**

**Машинное обучение в бизнесе**

---

**Уровень высшего образования:**

**МАГИСТРАТУРА**

---

**Направление подготовки:**

**38.04.01. ЭКОНОМИКА**

---

**Форма обучения:**

**ОЧНАЯ**

---

Рабочая программа рассмотрена и одобрена  
*Учебно-методической комиссией экономического факультета*  
(протокол № \_\_\_\_\_, дата)

Москва 2020

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки магистратуры 38.04.01. Экономика

ОС МГУ утвержден решением Ученого совета МГУ имени М.В.Ломоносова от 28 декабря 2020 года, протокол №7

Год (годы) приема на обучение: 2021 и последующие

# 1. Место и статус дисциплины в структуре основной профессиональной образовательной программы подготовки магистра

Статус дисциплины: *по выбору*

Триместр: 4

## 2. Входные требования (реквизиты) для освоения дисциплины

Для успешного освоения данного курса требуются знания и умения, полученные в следующих дисциплинах:

- Английский язык
- Эконометрика (продвинутый уровень)
- Машинное обучение и анализ данных-1

## 3. Планируемые результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	Индикаторы (показатели) достижения компетенций	Планируемые результаты обучения по дисциплине (модулю), сопряженные с компетенциями
МПК-4. Способен видеть логические связи в системе собранной, обработанной и проанализированной информации, и на основании этого разрабатывать рекомендации для лиц, принимающих решения на микро- и макро уровне, или бизнес-решения	МПК-4.И-1. Анализирует и систематизирует собранные данные, разрабатывает рекомендации для лиц, принимающих решения на основе собранных данных	МПК-4.И-1.У-1. Умеет строить логические взаимосвязи в системе собранной информации  МПК-4.И-1.У-2. Умеет разрабатывать рекомендации для лиц, принимающих управленческие решения как сфере государственного управления на разных уровнях, так и в сфере бизнеса
ОПК-5. Способен использовать современные информационные технологии и программные средства при решении профессиональных задач	ОПК-5.И-1. Применяет общие или специализированные пакеты прикладных программ, предназначенные для обработки, визуализации и анализа данных, в том числе эконометрического анализа и имитационного моделирования	ОПК-5.И-1.У-1. Умеет применять общие или специализированные пакеты прикладных программ (MS Excel, Eviews, Stata, SPSS, AnyLogic, Tableau и др.) или языки программирования (R, Python и др.), предназначенные для обработки, визуализации и анализа данных, в том числе эконометрического анализа и имитационного моделирования в, в соответствии со своими профессиональными задачами

#### 4. Объем дисциплины по видам занятий

Объем дисциплины составляет 3 зачетных единицы: 108 академических часов, в том числе 52 академических часа составляет контактная работа с преподавателем, 56 академических часов составляет самостоятельная работа магистранта.

#### 5. Формат обучения

очный, с использованием обучающей среды Op.Econ (при необходимости допускается применение дистанционных образовательных технологий).

#### 6. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий

Наименование и краткое содержание разделов и тем дисциплины (модуля),  Форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе							
		Контактная работа (работа во взаимодействии с преподавателем) <i>Виды контактной работы, часы</i>				Самостоятельная работа обучающегося  <i>Виды самостоятельной работы, часы</i>			
		Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Всего	Изучение материалов по курсу, чтение статей	Подготовка практических домашних заданий	Подготовка проекта	Всего
Тема 1. Байесовская статистика в бизнесе и принципы моделирования	32	8	8	-	16	8	4	4	16
Тема 2. Байесовские ABCDE-тесты и восстановление параметров	20	4	4	-	8	4	4	4	12
Тема 3. Отладка Байесовского вывода	20	4	4	-	8	4	4	4	12
Тема 4. Временные ряды и Гауссовские	32	8	8	-	16	8	4	4	16

процессы									
Промежуточная аттестация	4								
Защита проекта									
<b>Итого</b>	108								

## Краткое содержание тем дисциплины

### Тема 1. Байесовская статистика в бизнесе и принципы моделирования

#### Содержание темы:

Байесовский подход в статистике против частотного подхода классической эконометрики. Формула Байеса. Принцип достаточности и принцип правдоподобия. Примеры задач, решаемых в рамках байесовского подхода. введение в PyMC: Настройка окружения. Базовая модель. Параметризация, выбор априорного распределения. Проверка априорного распределения. НМС для практики. Разбор кейса по составлению модели. Иерархичность. Матрица адаптации. Проверка априорного распределения (практика). Освоение сервиса Yandex DataSphere, выполнение практической работы в облаке.

#### Основная литература:

1. Gelman et al. (2020) Bayesian workflow. Электронный ресурс, URL: <https://arxiv.org/abs/2011.01808>

#### Дополнительная литература:

1. <https://www.pymc.io/projects/docs/en/latest/installation.html> — Руководство по PyMC, рекомендовано к самостоятельному изучению
2. <https://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/> — Введение в байесовские методы и вероятностное программирование
3. <https://www.pymc.io/projects/docs/en/stable/learn.html> — Ресурс об использовании PyMC для Байесовского моделирования

### Тема 2. Байесовские ABCDE-тесты и восстановление параметров

#### Содержание темы:

Байесовские ABCDE-тесты. Восстановление параметров (Parameter Recovery): в каких случаях необходимо применять, алгоритм. Введение в Dask. Эффективный Xarray

*Основная литература:*

1. Gelman et al. (2020) Bayesian workflow. Электронный ресурс, URL: <https://arxiv.org/abs/2011.01808>
2. Dube, Jean-Pierre, Gunter Hitsch and Peter Rossi (2012), State dependence and alternative explanations for consumer inertia, RAND Journal of Economics Vol. 41, No. 3, Autumn 2010 pp. 417445
3. Martin O. Bayesian Analysis with Python - Second Edition. 2018. – 356 pp. ISBN 9781789341652 <https://www.packtpub.com/product/bayesian-analysis-with-python/9781789341652>

*Дополнительная литература:*

1. <https://www.pymc.io/projects/docs/en/latest/installation.html> — Руководство по PyMC, рекомендовано к самостоятельному изучению
2. <https://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/> — Введение в байесовские методы и вероятностное программирование
3. <https://www.pymc.io/projects/docs/en/stable/learn.html> — Ресурс об использовании PyMC для Байесовского моделирования

**Тема 3. Отладка Байесовского вывода**

*Содержание темы:*

Вычислительные проблемы НМС: обзор проблем на практике, обзор инструментов для решения проблем. Практика отладки НМС. Разбор примера с проблемами. Последовательность действий при дебаге НМС. Выполнение практической работы в облаке

*Основная литература:*

1. Gelman et al. (2020) Bayesian workflow. Электронный ресурс, URL: <https://arxiv.org/abs/2011.01808>
2. Martin O. Bayesian Analysis with Python - Second Edition. 2018. – 356 pp. ISBN 9781789341652 <https://www.packtpub.com/product/bayesian-analysis-with-python/9781789341652>

*Дополнительная литература:*

1. Salvatier J, et al. (2016) Probabilistic programming in Python using PyMC3 Электронный ресурс, URL: [https://www.pymc.io/projects/docs/en/latest/learn/core\\_notebooks/pymc\\_overview.html](https://www.pymc.io/projects/docs/en/latest/learn/core_notebooks/pymc_overview.html)
2. <https://www.pymc.io/projects/docs/en/latest/installation.html> — Руководство по PyMC, рекомендовано к самостоятельному изучению
3. <https://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/> — Введение в байесовские методы и вероятностное программирование
4. <https://www.pymc.io/projects/docs/en/stable/learn.html> — Ресурс об использовании PyMC для Байесовского моделирования

## Тема 4. Временные ряды и Гауссовские процессы

### Содержание темы:

Введение в Гауссовские процессы. Основные гиперпараметры. Разбор примера. Prior predictive для гауссовских процессов. Эффективная параметризация. Учет сезонности и трендов. Учет прочих факторов. Кросс-валидация. Разбор кейса по временной ряд: Модель без тренда, модель с трендом, модель с трендом и сезонностью. Сравнение моделей через LOO. Построение прогноза. Выполнение практической работы в облаке.

### Основная литература:

1. Gelman et al. (2020) Bayesian workflow. Электронный ресурс, URL: <https://arxiv.org/abs/2011.01808>
2. Martin O. Bayesian Analysis with Python - Second Edition. 2018. – 356 pp. ISBN 9781789341652 <https://www.packtpub.com/product/bayesian-analysis-with-python/9781789341652>

### Дополнительная литература:

1. <https://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/> — Введение в байесовские методы и вероятностное программирование
2. <https://www.pymc.io/projects/docs/en/stable/learn.html> — Ресурс об использовании PyMC для Байесовского моделирования

## 7. Фонд оценочных средств для оценивания результатов обучения по дисциплине

### 7.1. Примеры оценочных средств:

Результаты обучения по дисциплине	Виды оценочных средств
МПК-4.И-1.У-1. Умеет строить логические взаимосвязи в системе собранной информации	Домашние задания Проект
МПК-4.И-1.У-2. Умеет разрабатывать рекомендации для лиц, принимающих управленческие решения как в сфере государственного управления на разных уровнях, так и в сфере бизнеса	

ОПК-5.И-1.У-1. Умеет применять общие или специализированные пакеты прикладных программ (MS Excel, Eviews, Stata, SPSS, AnyLogic, Tableau и др.) или языки программирования (R, Python и др.), предназначенные для обработки, визуализации и анализа данных, в том числе эконометрического анализа и имитационного моделирования в, в соответствии со своими профессиональными задачами	Домашние задания Проект
--	----------------------------

## 7.2. Критерии оценивания (баллы) по дисциплине:

Виды оценочных средств	Баллы
Домашние задания	90
Проект	30
Экзамен в виде защиты проекта	30
<b>Итого</b>	<b>150</b>

## 7.3. Оценка по дисциплине выставляется, исходя из следующих критериев:

Оценка	Минимальное количество баллов	Максимальное количество баллов
<i>Отлично</i>	127,5	150,0
<i>Хорошо</i>	97,5	127,0
<i>Удовлетворительно</i>	60,0	97,0
<i>Неудовлетворительно</i>	0,0	59,5

**Примечание:** в случае, если магистрант за триместр набирает менее 20% баллов от максимального количества по дисциплине, то уже на промежуточном контроле (и далее на пересдачах) действует следующее правило сдачи: «магистрант может получить только оценку «Удовлетворительно», и только если получит за промежуточный контроль, включающий весь материал дисциплины, не менее, чем 85% от баллов за промежуточный контроль».

## 7.4. Типовые задания и иные материалы, необходимые для оценки результатов обучения:

— *Пример задания для проекта:*

Проект представляет собой небольшой текст (до 15 стр.), содержащий постановку проблемы, краткий обзор литературы, разбор теоретической модели (или моделей) с предпосылками и выводами, а также подбор примеров из эмпирических статей, иллюстрирующих выводы модели и собственные расчёты. Необходим подбор данных и их описание. Проект выполняется в группах по 2-3 человека и сдаётся в виде текста в формате pdf, содержащего необходимые таблицы и графики, а также Jupiter Notebook. Расчёты выполняются с



использованием сервисов Yandex DataSphere.

После сдачи проектов проходит перекрёстное рецензирование. Предполагается рецензия на проект коллег и устное выступление с ней (вопросы, замечания, комментарии).

— *Примеры домашних заданий*

*Пример теоретического задания:*

Вопрос 1.

Вам предлагается оценить долю посетителей сайта интернет-магазина, которые купят товар А. Чтобы это сделать, вы берёте случайную выборку размера  $n$  из посетителей сайта. Поскольку Вы обладаете ограниченным временем и ресурсами, Ваша выборка относительно мала. Предположим, что  $n=20$ . После этого Вы выясняете, что 6 человек из Вашей выборки покупают товар А. Затем Вы смотрите на данные за предшествующий месяц и выясняете, что 40% посетителей сайта купили товар А.

Как Вы можете использовать эту информацию, чтобы улучшить оценку доли посетителей, которые купят товар А?

Вопрос 2.

В магазине компании N каждый третий день раскупают весь товар. Аналитик пытается предсказать, раскупят ли весь товар на следующий день. 75% дней с полной распродажей и 3 из 5 дней с остатками товара были верно предсказаны аналитиком в прошлый раз. Если в этот раз аналитик предсказывает полную распродажу товара, то какова вероятность, что завтра действительно весь товар будет распродан?

*Пример практического задания:*

Задание выполняется студентом индивидуально, сдаётся на онэкон в виде 2 файлов – оформленного связного текста с ответами на вопросы, таблицами и графиками (если они необходимы по заданию) в формате pdf и кода в Jupiter Notebook. В первой строчке кода в качестве комментария после решетки напишите, пожалуйста, имя автора.

В этом задании вам предлагается дать оценку результатов проведения АБ-теста, который проводил аналитик в компании N. Используемые данные находятся в файле price.csv.

- Выберите результирующую метрику.
- Постойте эмпирическое распределение для показателей, проверьте с помощью соответствующих тестов, что оно статистически отличается от нормального. Объясните, почему вы используете именно эти тесты?
- С учётом полученного результата проведите тесты на сравнение средних в 2 группах для выбранной метрики.
- Напишите краткий отчёт с итоговым решением для ценовой политики компании N.

— **Методические рекомендации и требования к выполнению заданий:**

— *Подготовка домашних заданий:*

Для выполнения задания необходимо изучить материалы семинаров по теме, а также главы из основной литературы. Задание выполняется индивидуально, сдаётся на онэкон в виде pdf-файла с отчётом (связный, оформленный текст с нужными пояснениями таблицами,

графиками) и кода, написанного в Jupiter Notebook с использованием сервисов Yandex DataSphere. В отчёте наряду с содержанием оценивается качество оформления: таблицы, графики, выводы. Задание необходимо выполнять по вариантам. В случае нарушения академической этики (списанного текста либо кода) баллы за задание обнуляются.

— *Подготовка проекта*

Для выполнения задания необходимо изучить материалы семинаров по теме, а также главы из основной литературы. Задание выполняется в группах по 2-3 человека. Тему проекта необходимо согласовать с преподавателем до утверждённого дедлайна. В случае нарушения академической этики (списанного текста) баллы за задание обнуляются.

— *Подготовка к защите проекта (экзамену)*

Для подготовки к защите проекта (экзамену) необходимо подготовить презентацию до 20 слайдов на 15 минут, а также ответы на вопросы из рецензии.

## 8. Ресурсное обеспечение

### 8.1. Перечень основной и дополнительной литературы

#### Основная литература:

4. Gelman et al. (2020) Bayesian workflow. Электронный ресурс, URL: <https://arxiv.org/abs/2011.01808>
5. Dube, Jean-Pierre, Gunter Hitsch and Peter Rossi (2012), State dependence and alternative explanations for consumer inertia, RAND Journal of Economics Vol. 41, No. 3, Autumn 2010 pp. 417-445
6. Martin O. Bayesian Analysis with Python - Second Edition. 2018. – 356 pp. ISBN 9781789341652 <https://www.packtpub.com/product/bayesian-analysis-with-python/9781789341652>

#### Дополнительная литература:

1. Salvatier J, et al. (2016) Probabilistic programming in Python using PyMC3 Электронный ресурс, URL: [https://www.pymc.io/projects/docs/en/latest/learn/core\\_notebooks/pymc\\_overview.html](https://www.pymc.io/projects/docs/en/latest/learn/core_notebooks/pymc_overview.html)
2. Robert C. P. et al. The Bayesian choice: from decision-theoretic foundations to computational implementation. – New York : Springer, 2007. – Т. 2.
3. Rossi, P.E., G.M. Allenby and R. McCulloch (2005) Bayesian Statistics and Marketing, John Wiley & Sons, Ltd.

## 8.2. Перечень лицензионного программного обеспечения

— Используемые сервисы: Yandex Cloud, Yandex DataSphere.

## 8.3. Перечень профессиональных баз данных и информационных справочных систем

— *Не требуется*

## 8.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (при необходимости)

4. <https://www.pymc.io/projects/docs/en/latest/installation.html> — Руководство по PyMC, рекомендовано к самостоятельному изучению
5. <https://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/> — Введение в байесовские методы и вероятностное программирование
6. <https://www.pymc.io/projects/docs/en/stable/learn.html> — Ресурс об использовании PyMC для Байесовского моделирования
7. <https://www.nber.org/lecture/summer-institute-2007-methods-lecture-bayesian-inference> — Imbens G., J.M. Wooldridge (2007) Lecture Notes for NBER Summer School: Bayesian Inference

## 8.5. Описание материально-технической базы

Для организации занятий по дисциплине необходимы следующие технические средства обучения: компьютерный класс с установленной, проектором, доской, маркеры. Для подключения к сервисам Yandex DataSphere студентам необходимо завести электронный ящик на @yandex.ru

## 9. Язык преподавания:

*Английский*

## 10. Преподаватель (преподаватели):

*Кочуров Максим Вадимович, Партнер PyMC Labs (внешний преподаватель)*

## 11. Разработчики программы:

*Кочуров Максим Вадимович, Партнер РумС Labs (внешний преподаватель)*