

**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет экономических наук
Школа финансов

Рабочая программа дисциплины «IT для финансистов»

для образовательной программы «Финансовые рынки и финансовые институты»
направления 38.04.08 «Финансы и кредит»
подготовки магистра

Разработчики программы:

Гаращук Г.В., ggarashchuk@hse.ru

Науменко В.В., v.v.naumenko@gmail.com

Булгаков А.Л., z3900207@mail.ru

Одобрена на заседании Школы финансов

«__» _____ 201__ г.

Руководитель Школы финансов

И.В. Ивашковская _____

Утверждена Академическим советом образовательной программы «Финансовые
рынки и финансовые институты»

«__» _____ 201__ г., № протокола _____

Академический руководитель образовательной программы

Т.В. Теплова _____

Москва, 2017

*Настоящая программа не может быть использована другими подразделениями университета и
другими вузами без разрешения подразделения-разработчика программы.*

1. Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов направления подготовки для направления 38.04.08 «Финансы и кредит» подготовки магистра для магистерской программы «Финансовые рынки и финансовые институты» изучающих дисциплину «IT для финансистов».

Программа разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ;
- Образовательной программы для направления 38.04.08 «Финансы и кредит» подготовки магистра для магистерской программы «Финансовые рынки и финансовые институты»
- Рабочим учебным планом университета по направлению подготовки для направления 38.04.08 «Финансы и кредит» подготовки магистра для магистерской программы «Финансовые рынки и финансовые институты», утвержденным в 2017г.

2. Цели освоения дисциплины

Целью освоения дисциплины «IT для финансистов» является освоение студентами практической Data Science (науки о данных), то есть современных информационных технологий для извлечения, преобразования и анализа финансовых и экономических данных.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- уметь разрабатывать схему базы данных на основе анализа предметной области;
- знать нормальные формы реляционных баз данных и уметь производить нормализацию имеющихся баз данных;
- знать основы языка запросов SQL;
- знать основы OLAP;
- уметь работать в R/RStudio с базами данных, развернутыми в Microsoft Access / Microsoft SQL Server / Oracle Database;
- уметь пользоваться возможностями OLAP в Microsoft Excel.

4. Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы			Самостоятельная работа
			Лекции	Семинары	Практические занятия	
1	Введение в Data Science. Основные этапы проекта Data Science. Методика поиска и извлечения данных из открытых источников (сеть Интернет).	14	2	4		8
2	Основы программирования в R. Среда RStudio. Основные структуры и понятия.	12		4		8
3	Продвинутая визуализация данных в R. Реализация концепции Reproducible Research посредством RMarkdown	12		4		8
4	Предобработка данных и реализация линейных статистических моделей в R.	12		4		8
5	Построение торговых стратегий, back-tetsing и диагностика торговых стратегий в R.	20		8		16
6	Анализ финансовых данных. Математические модели выбора оптимального торгового решения	4	2	2		-
7	Продвинутые модели прогнозирования и реализация торговых стратегий в R.	6		6		13

8	Основные понятия баз данных.	26		6		13
9	Извлечение данных из баз данных и языков запросов SQL.	24		6		13
10	Обработка данных с помощью технологии OLAP	24		6		13
11	Лабораторная работа № 1. Анализ данных и построение стратегий на фондовом рынке.	12		2		10
12	Лабораторная работа № 2. Анализ данных и построение стратегий на рынке опционов.	12		2		10
13	Лабораторная работа № 3. Анализ данных и построение стратегий на валютном рынке.	12		2		10
	Итого	190	4	56		130

5. Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Кафедра
		1	2	3	4	
Текущий (неделя)	Работа в аудитории			*		
	Домашнее задание			*		
Итоговый	Экзамен			*		

Параметры **
Письменный экзамен

6. Критерии оценки знаний, навыков

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

7. Содержание дисциплины

Раздел 1. Введение в Data Science. Основные этапы проекта Data Science. Методика поиска, извлечения и преобразования данных

Вводится понятие проекта Data Science. Подробно рассматриваются все этапы проекта Data Science и соответствующее программное обеспечение. Подробно изучаются нестандартные методы поиска и извлечения данных из открытых источников в сети Интернет, загрузка и первичная обработка данных, автоматизация мониторинга данных в сети Интернет (персональные поисковые роботы).

Раздел 2. Основы программирования в R. Среда RStudio. Основные структуры и понятия.

Осуществляется знакомство с языком R и средой программирования RStudio. Вводятся основные структуры и понятия языка R: переменные, типы данных, векторизованные вычисления, функции и др.

Раздел 3. Продвинутое визуализация данных в R. Реализация концепции Reproducible Research посредством RMarkdown.

Изучаются возможности визуализации данных с помощью пакета ggplot2 и др. Подготовка отчета об исследовании через инструмент RMarkdown.

Раздел 4. Предобработка данных и реализация линейных статистических моделей в R.

Рассматриваются основные проблемы предобработки данных. Выгрузка данных из открытых источников. Использование инструментария R (функции семейства apply, пакет dplyr, magrittr и др.) для предобработки данных. Дисперсионный и регрессионный (линейная регрессия) анализ данных посредством R.

Раздел 5. Построение торговых стратегий, back-testing и диагностика торговых стратегий в R.

Базовые идеи торговых стратегий. Изучение элементов торговой стратегии (сигнал, исполнение, управление риском и т.д.). Осуществление корректного тестирования стратегий. Изучение методов диагностики и отбора стратегий: ROC, график накопленной доходности, тесты на значимость и др.

Раздел 6. Продвинутые модели прогнозирования и реализация торговых стратегий в R.

Рассматриваются различные параметрические и непараметрические модели для анализа финансовых рынков и построения торговых стратегий: стохастические дифференциальные уравнения, динамические системы, докомпозиция и методы главных и независимых компонент, реконструкция расчетных цен и риск-нейтральной плотности вероятности на рынке опционов.

Раздел 7. Основные понятия баз данных.

Базы данных: физическая и логическая структура. Типы баз данных. Реляционная модель баз данных. Системы управления базами данных (СУБД). Нормализация баз данных. Объекты базы данных: таблицы, индексы, представления, последовательности.

Раздел 8. Извлечение данных из баз данных и язык запросов SQL.

Подключение к базе данных из R. Язык запросов SQL: язык определения данных (DDL) и язык манипулирования данными (DML). Операторы манипулирования данными в SQL: выборка, добавление, обновление и удаление данных. Групповые функции. Объединение данных. Вложенные запросы. Иерархические запросы.

Раздел 9. Обработка данных с помощью технологии OLAP.

Введение в многомерный анализ данных. Типы OLAP. Аналитическая обработка данных в реальном времени: OLAP-кубы. Операции над OLAP-кубом. Средства анализа и моделирования данных в MS Excel: сводные таблицы и диаграммы, настройка Power Pivot.

8. Образовательные технологии

- Методические рекомендации преподавателю
- Методические указания студентам

9. Оценочные средства для текущего контроля и аттестации студента

Тематика заданий текущего контроля

Домашние задания по тематикам Разделов 1-4

Лабораторная работа № 1. Анализ данных и построение стратегий на фондовом рынке

Лабораторная работа № 2. Анализ данных и построение стратегий на рынке опционов

Лабораторная работа № 3 Анализ данных и построение стратегий на валютном рынке

Вопросы для оценки качества освоения дисциплины

Примерный перечень вопросов к зачету (экзамену) по всему курсу или к каждому промежуточному и итоговому контролю для самопроверки студентов.

- Синтаксис и структура скрипта в среде R/RStudio.
- Средства R/RStudio для визуализации данных.
- Стандартные средства оценки качества эконометрической модели функции $\text{lm}()$.
- Логистическая регрессия.
- Деревья классификации и регрессии.
- Загрузить из сети Интернет набор данных по заданной преподавателем проблеме и ввести его в R/RStudio.
- Загрузить данные из базы данных в R/RStudio.
- Преобразовать набор данных R/RStudio по заданию преподавателя.
- Модифицировать набор данных R/RStudio, содержащий пропущенные значения.
- Модифицировать набор данных R/RStudio, содержащий выбросы.
- Создать в R/RStudio линейную регрессионную модель и объяснить структуру созданного модельного объекта.
- Провести в R/RStudio детальную диагностику линейной регрессионной модели.
- Осуществить back-testing торговой стратегии.
- Оформить результаты исследования через инструмент RMarkdown.

10. Порядок формирования оценок по дисциплине

Преподаватель оценивает самостоятельную работу студентов. Оценки за самостоятельную работу студента преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за самостоятельную работу определяется перед промежуточным или итоговым контролем – $O_{\text{сам. работа}}$.

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

$$O_{\text{накопленная}} = 0.4 * O_{\text{дз}} + 0.6 * O_{\text{лаб.}}$$

где $O_{\text{дз}}$ – средняя оценка за домашние задания;

$O_{\text{лаб.}}$ – средняя оценка за лабораторные работы.

Результирующая оценка за дисциплину рассчитывается следующим образом:

$$O_{\text{итоговая}} = 0.5 * O_{\text{накопленная}} + 0.5 * O_{\text{экзамен}}$$

Экзамен представляет собой защиту проекта, заключающегося в анализе финансовых данных, построении, диагностике и тестировании торговой стратегии посредством языка R. Представление проекта должно быть выполнено через RMarkdown с использованием инструментов визуализации. Также должен быть предоставлен код с комментариями. В рамках экзамена могут быть заданы дополнительные вопросы по курсу.

12. Учебно-методическое и информационное обеспечение дисциплины

12.1. Базовый учебник

Nina Zumel, John Mount. Practical Data Science with R. Manning Publications Co., 2014.
Richard Cotton. Learning R: A Step-by-Step Function Guide to Data Analysis. O'Reilly, 2013.
Hadley Wickham. Advanced R (Chapman & Hall/CRC The R Series). CRC Press.2015

12.2. Основная литература

Diethelm Würtz, Tobias Setz, Yohan Chalabi, Longhow Lam, Andrew Ellis. Basic R for Finance. Rmetrics Association & Finance Online Publishing, Zurich, 2010.

Frontini, M., Tagliani, A.: Entropy-convergence in Stieltjes and Hamburger moment problem. Applied Mathematics and Computation, Volume 88, 1997, pp. 39-51

Golyandona, N., Korobeynikov, A.: Basic Singular Spectrum Analysis and Forecasting with R. Preprint submitted to Computational Statistics & Data Analysis, January, 2013

Golyandina, N., Nekrutkin, V., Zhigljavsky, A.: Analysis of Time Series Structure: SSA and Related Techniques. London: Chapman & Hall/CRC, 2001, p. 305

Breeden, D.T., Litzenberger, R.H., Prices of State-contingent Claims Implicit in Option Prices // Journal of Business, 1978, 51(4), pp. 621-651

Figlewski, St., Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio (July 30, 2008). Oxford, UK: Oxford University Press, 2008

Aapo Hyvärinen, Erkki Oj. Independent Component Analysis: Algorithms and Applications. Neural Networks, 13(4-5):411-430, 2000

12.3. Дополнительная литература

- R Project Documentation (www.r-project.org/doc/bib/R-books.html).
- Kurt Hornik. Frequently Asked Questions on R (cran.r-project.org/doc/FAQ/R-FAQ.html)
- R: Анализ и визуализация данных (<http://r-analytics.blogspot.ru/>)
- Сергей Мастицкий, Владимир Шитиков. Статистический анализ и визуализация данных с помощью R. ДМК Пресс, 2015.
- Christopher Gandrud, Reproducible Research with R and R Studio, Second Edition, Chapman and Hall/CRC, 2015.
- Robert I. Kabacoff. R in Action, Second Edition. Data Analysis and Graphics with R. Manning Publications Co., 2015.
- Norman Matloff. The Art of R Programming. A Tour of Statistical Software Design. No Starch Press, 2011.
- Гельфанд И.М., Фомин С.В. Вариационное исчисление. М.: Государственное издательство физико-математической литературы, 1961.
- Шеннон К. Работы по теории информации и кибернетике. — М.: Изд.иностр.лит., 2002

12.4. Справочники, словари, энциклопедии

12.5. Программные средства

- R/RStudio
- Microsoft Access, H2, SQLite, SQuireL SQL, SQL Screwdriver

- Microsoft Excel 2016, Power BI
- MikTeX

12.6. Дистанционная поддержка дисциплины

13. Материально-техническое обеспечение дисциплины

Проектор, компьютерный класс.