

**Санкт-Петербургский филиал федерального государственного
автономного образовательного учреждения высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет Санкт-Петербургская школа экономики и менеджмента
Национального исследовательского университета
«Высшая школа экономики»

Департамент прикладной математики и бизнес-информатики

**Рабочая программа дисциплины
Корпоративный анализ данных**

для образовательной программы «Анализ больших данных в бизнесе, экономике и обществе»
направления подготовки 01.04.02 «Прикладная математика и информатика»
уровень магистратура

Разработчик(и) программы

Мусабилов И.Л., MA (Digital Media and Information Society), магистр по направлению
«Информатика и вычислительная техника», преподаватель департамента социологии,
ilya@musabirov.info

А.В.Сироткин, канд. физ.-мат. наук, доцент департамента прикладной математики и бизнес-
информатики, avsirotkin@hse.ru

Согласована менеджером ОП «Анализ больших данных в бизнесе, экономике и обществе»

«30» августа 2016г.

Е.С. Авдонина _____

Утверждена Академическим руководителем образовательной программы

А.В. Сироткин _____

«30» августа 2016г.

Санкт-Петербург, 2016

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



1 Область применения и нормативные ссылки

Настоящая рабочая программа дисциплины устанавливает минимальные требования к знаниям и умениям студента, а также определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Корпоративный анализ данных», учебных ассистентов и студентов направления подготовки 01.04.02 «Прикладная математика и информатика», обучающихся по образовательной программе «Анализ больших данных в бизнесе, экономике и обществе».

Рабочая программа дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ
<http://www.hse.ru/data/2016/11/02/1111123560/01.04.02%20Прикладная%20математика%20и%20информатика.pdf>;
- Образовательной программой «Анализ больших данных в бизнесе, экономике и обществе», направление подготовки 01.04.02 «Прикладная математика и информатика»;
- Объединенным учебным планом университета по образовательной программ «Анализ больших данных в бизнесе, экономике и обществе».

2 Цели освоения дисциплины

Целями освоения дисциплины «Корпоративный анализ данных» является формирование навыков работы с анализом данных как процессом. Дисциплина сочетает в себе как непосредственно работу с данными, изучение предиктивных моделей, так и изучение процессов внедрения решений, на основе анализа в бизнес-процессы.

Часть лекций в данном курсе проводится приглашенными аналитиками имеющими богатый опыт работы в реальной индустрии.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Уровни формирования компетенций:

РБ - ресурсная база, в основном теоретические и предметные основы (знания, умения)

СД - способы деятельности, составляющие практическое ядро данной компетенции

МЦ - мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
<i>Системные компетенции</i>					
Способен принимать управленческие решения и готов нести за них	СК-5	РБ, СД	Способен выбирать направление проекта с учетом имеющихся ресурсов, предлагать путь	Групповые выступления на семинарах, подготовка	Домашнее задание



ответственность			достижения цели и следовать ему	командного проекта	
<i>Профессиональные компетенции</i>					
<i>А) Социально-личностные компетенции</i>					
Способен к осознанному выбору стратегий межличностного взаимодействия	ПК-5	СД	Способен эффективно работать в команде, совместно добиваясь решения поставленной задачи.	Групповые выступления на семинарах, подготовка командного проекта	Домашнее задание.
<i>Б) Инструментальные компетенции</i>					
Способен описывать проблемы и ситуации профессиональной деятельности, используя язык и аппарат прикладной математики при решении междисциплинарных проблем.	ПК-14	СД	Способен формально описывать задачи, возникающие в бизнес-аналитике, и сводить их к математическим или технологическим задачам.	Семинарские занятия, подготовка командного проекта	Домашнее задание, аудиторная работа, экзамен
Способен создавать, описывать и ответственно контролировать выполнение технологических требований и нормативных документов в профессиональной деятельности.	ПК-15	СД	Способен подготовить и согласовать ТЗ на разрабатываемый совместно программный продукт/модель. Способен сам следовать согласованному ТЗ и следить за выполнением ТЗ другими участниками.	Семинарские занятия, подготовка командного проекта	Домашнее задание
Способен строить и решать математические модели в соответствии с направлением подготовки и специализации	ПК-17	СД	Умеет сводить прикладные задачи к задачам регрессии, классификации и задачам предиктивного моделирования.	Лекции, семинарские занятия, подготовка командного проекта	Домашнее задание, аудиторная работа, экзамен

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к циклу дисциплин по выбору студента магистерской программы «Анализ больших данных в бизнесе, экономике и обществе».

Изучение данной дисциплины базируется на следующих дисциплинах магистерской программы: «Алгоритмы и структуры данных», «Математические основы анализа данных»

Для освоения учебной дисциплины, студенты должны владеть следующими знаниями и компетенциями



- способен принимать управленческие решения и готов нести за них ответственность;
- способен к осознанному выбору стратегий межличностного взаимодействия;
- способен описывать проблемы и ситуации профессиональной деятельности, используя язык и аппарат прикладной математики при решении междисциплинарных проблем;
- способен создавать, описывать и ответственно контролировать выполнение технологических требований и нормативных документов в профессиональной деятельности
- способен строить и решать математические модели в соответствии с направлением подготовки и специализации.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- «Современные методы анализа данных»,

5 Тематический план учебной дисциплины

ОБЪЕМ ДИСЦИПЛИНЫ - 4 зачетные единицы.

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Семинары	
1	Введение в корпоративную аналитику данных	34	4	4	26
2	Инфраструктура анализа данных	72	6	6	60
3	Организация аналитики в компании	46	6	6	34
ИТОГО		152	16	16	120

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	2 Модуль		Параметры
		Ноябрь	Декабрь	
Текущий	Домашнее задание		*	Студенты в группах по 3-4 человека готовят приложение или сервис, основанный на данных
	Аудиторная работа	*	*	Выполнение практических задач связанных с анализом. Ответы по теме занятия и участие в дискуссии.
Итоговый	Экзамен		*	Письменный экзамен - 60 минут

7 Критерии оценки знаний, навыков

Критерии оценки домашнего задания:

Домашнее задание проводится в виде сдачи программы или сервиса на платформе R/Shiny/ShinyApps, языке Python, платформе Azure ML или другой платформе, по согласованию с



преподавателем. Работа оценивается “зачтено”/ “не зачтено”, при этом в случае оценки “зачтено” в формулу оценки в качестве балла за домашнюю работу проставляется 10, в противном случае -- 0.

Критериями оценки являются:

- корректность запросов к базе данных, определения функционала, реализуемого на стороне БД и в программном коде
- корректность применения алгоритмов машинного обучения
- оригинальность идеи

Нетривиальные и исключительные решения, разработанные студентами, получают 0,5 балла к результирующей оценке за курс.

Критерии оценки аудиторной работы:

В процессе семинаров преподаватель оценивает активность студентов, уровень их участия в дискуссии. Отдельно оценивается выполнение заданий в классе, связанных с подготовкой скриптов по анализу данных.

Участие в дискуссии оценивается максимум в 4 балла. 4 балла студент получает если неоднократно участвует в диалоге и дает обоснованные ответы. 2 балла за дискуссии студент получает за активное обсуждение по крайней мере на одном семинаре. 0 баллов выставляется, если студент не участвует в обсуждении в течении всего модуля.

Практическое выполнение заданий оценивается преподавателем в 6 баллов. 6 баллов выставляется за выполнение всех текущих заданий и демонстрации результатов преподавателю. 3 балла ставится за выполнения большей части заданий. 0 баллов выставляется, если студент не выполняет задания в классе и не может продемонстрировать результаты своей работы преподавателю.

Итоговая оценка за аудиторную работу складывается из двух приведенных выше.

Максимальная оценка за аудиторную работу – **10 баллов.**

Критерии оценки за экзамен:

Экзамен состоит из набора закрытых и открытых вопросов, освещающих основные темы курса. Вопросы могут быть теоретическими или в форме задачи по пройденному материалу.

Экзамен проводится в письменной форме. Максимальные баллы за каждый вопрос указываются в экзаменационном задании.

Максимальная оценка за экзамен – **10 баллов.**

8 Содержание дисциплины

Раздел 1. Введение в корпоративную аналитику данных

Аналитическое мышление. Бизнес-проблемы и наука о данных. Введение в предиктивное моделирование. Дата-продукты.

Литература по разделу

- Cavanillas, José María, Edward Curry, and Wolfgang Wahlster, eds. 2016. *New Horizons for a Data-Driven Economy*. Cham: Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-21569-3>: Chapters 5, 8
- Davenport, Thomas H. 2014. “What Businesses Can Learn from Sports Analytics.” *MIT Sloan Management Review* 55 (4): 10.
- Data Jujitsu: The art of turning data into product. <http://radar.oreilly.com/2012/07/data-jujitsu.html>
- Designing great data products. The Drivetrain Approach: A four-step process for building data products. <https://www.oreilly.com/ideas/drivetrain-approach-data-products>



Раздел 2. Инфраструктура анализа данных.

Базы данных и хранилища данных. Функциональные классы аналитических систем. Системы оптимизации. Экспертные системы. Системы машинного обучения. Операционная бизнес-аналитика. Аналитическая отчетность. ERP-системы. Облачные решения анализ данных.

Литература по разделу

- Cavanillas, José María, Edward Curry, and Wolfgang Wahlster, eds. 2016. *New Horizons for a Data-Driven Economy*. Cham: Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-21569-3>: Chapters 4, 7
- Davenport, Thomas H. 2016. "Rise of the Strategy Machines." *MIT Sloan Management Review* 58 (1): 29.
- Verhoef, Peter C., Penny N. Spring, Janny C. Hoekstra, and Peter SH Leeflang. 2003. "The Commercial Use of Segmentation and Predictive Modeling Techniques for Database Marketing in the Netherlands." *Decision Support Systems* 34 (4): 471–481.

Раздел 3. Организация аналитики в компании.

Проектная и процессная организация аналитики. Business Intelligence. Business Analytics. Enterprise Decision Management. Data Science и Big Data.

Литература по разделу

- Cavanillas, José María, Edward Curry, and Wolfgang Wahlster, eds. 2016. *New Horizons for a Data-Driven Economy*. Cham: Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-21569-3>: Chapter 15
- Davenport, Thomas H., and Stephan Kudyba. 2016. "Designing and Developing Analytics-Based Data Products." *MIT Sloan Management Review* 58 (1): 83.
- H. Davenport, Thomas. 2014. "How Strategists Use 'big Data' to Support Internal Business Decisions, Discovery and Production." *Strategy & Leadership* 42 (4): 45–50.

9 Образовательные технологии

9.1 Методические рекомендации преподавателю

Задача преподавателя в курсе с гостевыми лекторами -- не только служить модератором общения студентов с гостями, но и ставить гостевые лекции в более широкий, системный контекст изучаемого курса.

9.2 Методические указания студентам по освоению дисциплины

Самостоятельная работа студента включает:

- поиск и анализ информации по тематике курса;
- знакомство со специальной научной литературой по тематике курса;
- подготовку к семинарским занятиям.

9.3 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

- Analysing and Visualising Data with Power BI <https://www.edx.org/course/analyzing-visualizing-data-power-bi-microsoft-dat207x-3>



- Marketing Analytics: Competitive Analysis and Market Segmentation <https://www.edx.org/course/marketing-analytics-competitive-analysis-uc-berkeleyx-busadm466-3x>
- Knowledge Management and Big Data in Business <https://www.edx.org/course/knowledge-management-big-data-business-hkpolyux-ise101x-1>
- Big Data: from Data to Decisions <https://www.futurelearn.com/courses/big-data-decisions>

10 Оценочные средства для текущего контроля и аттестации студента

10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

Примеры проектов для домашнего задания:

1. Сервис рекомендации музыкальных композиций, на основе аккаунта из социальной сети.
2. Сервис отслеживания наиболее быстро распространяющихся новостей и прогноза их дальнейшей динамики.

Темы для обсуждения на семинаре:

1. Продукты, основанные на данных. Технологический стек, содержательные особенности, бизнес-модели
2. Инфраструктура анализа данных. Базы и хранилища данных, облачные сервисы, распределенная обработка данных
3. Корпоративный анализ данных как средство автоматизации скрытых оперативных решений
4. Карьерные треки и специализации в анализе данных

Примеры заданий на семинаре:

1. Выделите ТОП 10 Пользователей с максимальной репутацией
2. Найдите ТОП 10 постов с максимальным рейтингом
3. Определите 50 самых популярных тегов (Какие таблички нам для этого нужны?)
4. Вычислите суммарные scores для 50 самых популярных тегов.
5. Какие посты содержат тег R или тег Python?
6. Подготовьте скрипт выделяющий теги наиболее часто встречающиеся с тегами Python и R.
7. Постройте распределение по популярности для 1000 тегов.

10.2 Примеры заданий итогового контроля

Примеры теоретических вопросов экзамена:

1. Бизнес-проблемы и наука о данных. Формулировка бизнес-проблем. Решения, основанные на данных.
2. Проектная и процессная организация аналитики.
3. Business Intelligence. Business Analytics. Определение и соотношение понятий
4. Enterprise Decision Management. Суть концепции. Смысл управления решениями в организации
5. Data Science и Big Data. Определение и соотношение понятий с точки зрения бизнеса и инженерии
6. Базы данных и хранилища данных.
7. Функциональные классы аналитических систем.
8. Системы оптимизации. Экспертные системы. Системы машинного обучения.
9. Операционная бизнес-аналитика. Аналитическая отчетность. ERP-системы.



10. Облачные решения в области бизнес-аналитики
11. Инфраструктура анализа данных
12. Приложения и сервисы, основанные на данных
13. А/В тестирование и оптимизационные алгоритмы.

Примеры практических вопросов на экзамене:

1. Определите наиболее часто встречающийся объект в данном наборе данных.
2. Извлеките наиболее часто встречающуюся пару тегов.
3. Определите количество пользователей оставивших не более трех комментариев.

11 Порядок формирования оценок по дисциплине

Накопленная оценка по дисциплине рассчитывается с помощью взвешенной суммы оценок за отдельные формы текущего контроля знаний следующим образом:

$$O_{\text{накопленная}} = 0,7O_{\text{дз}} + 0,3O_{\text{ауд}}, \text{ где}$$

$O_{\text{дз}}$ - оценка знаний студента за домашнее задание;

$O_{\text{ауд}}$ – оценка знаний студента за аудиторную работу;

Результирующая оценка по дисциплине (которая идет в диплом) рассчитывается следующим образом:

$$O_{\text{результ}} = 0,8O_{\text{накопленная}} + 0,2O_{\text{экс}}, \text{ где}$$

$O_{\text{накопленная}}$ - накопленная оценка по дисциплине;

$O_{\text{экс}}$ - оценка за экзамен.

В формулу для $O_{\text{результ}}$ подставляются значения $O_{\text{накопленная}}$ и $O_{\text{экс}}$, округленные до ближайшего целого значения. $O_{\text{результ}}$ округляется до ближайшего целого значения.

По усмотрению ведущего преподавателя, если это не противоречит действующим документам на момент экзамена, при получении накопленной оценки 8 баллов и более, студент может быть освобожден от экзамена. В таком случае, с согласия студента, ему выставляется результирующая оценка, равная накопленной.

Студент не получает возможность пересдать низкие результаты за домашнюю работу а также при пропуске соответствующих им учебных часов.

При получении неудовлетворительной оценки $O_{\text{результ}}$ (значение после округления менее 4 баллов) выставляется оценка «НЕУДОВЛЕТВОРИТЕЛЬНО».

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Основная литература

Cavanillas, J.. New Horizons for a Data-Driven Economy [Electronic Resource] / José María Cavanillas, Edward Curry, Wolfgang Wahlster - Springer, 2016. – 303 p. - Authorized access: <http://link.springer.com/10.1007/978-3-319-21569-3> (Online Digital Library "Springer eBooks").



12.2 Дополнительная литература

- 1) Hastie T. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [Electronic Resource] / Trevor Hastie, Robert Tibshirani, Jerome Friedman. - Springer, 2009.- Authorized access: <http://link.springer.com/book/10.1007/978-0-387-84858-7> (Online Digital Library "Springer eBooks").
- 2) Siegel E. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Predictive Analytics [Electronic Resource] / Eric Siegel – Hoboken, N.J: Wiley, 2013. – 320 p. - Authorized access: <http://library.books24x7.com/toc.aspx?bookid=52840> (Online Digital Library "Books24x7").
- 3) Taylor J. Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics. Decision Management System [Electronic Resource] /James Taylor – NJ: IBM Press, 2012.- 316 p. - Authorized access: <http://library.books24x7.com/toc.aspx?bookid=45443> (Online Digital Library "Books24x7").
- 4) Blattberg R. Database Marketing: Analyzing and Managing Customers [Electronic Resource] / Robert C. Blattberg, Byung-Do Kim, Scott A. Neslin /- New York, NY: Springer, (International Series in Quantitative Marketing Vol. 18). 2008, 871 p. - Authorized access: <http://link.springer.com/10.1007/978-0-387-72579-6>. (Online Digital Library "Springer eBooks")/.

12.3 Справочники, словари, энциклопедии

-

12.4 Ресурсы информационно-телекоммуникационной сети «Интернет»

1. <https://azure.microsoft.com/ru-ru/services/machine-learning/>
2. <http://blog.revolutionanalytics.com/>
3. <http://www.tomdavenport.com/blogs-articles/>
4. <https://www.thoughtworks.com/insights/blog/introducing-agile-analytics>

12.5 Программные средства

PowerBI, RStudio

12.6 Информационные справочные системы

-

12.7 Дистанционная поддержка дисциплины

При необходимости возможна дистанционная поддержка.

13 Материально-техническое обеспечение дисциплины

При проведении лекционных и семинарских занятий преподавателем может использоваться компьютер и мультимедийный проектор.