

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

Факультет Санкт-Петербургская школа
физико-математических и компьютерных наук

основная образовательная программа «Анализ больших данных в бизнесе, экономике и
обществе»
направление подготовки 01.04.02 «Прикладная математика и информатика»
уровень магистратура, очная

Аннотации к рабочим программам дисциплин

СОГЛАСОВАНО:
Академический руководитель
образовательной программы:
_____ А.В. Сироткин
« 31 » августа 2017 г.

Санкт-Петербург, 2017 г.

Современные методы анализа данных

Тип дисциплины: обязательная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Математические основы анализа данных

Объем з.е.: 6

Описание курса: Целью освоения дисциплины «Современные методы анализа данных» является изучение основных аппаратов машинного обучения, эффективных алгоритмов обучения и применения обученных моделей, основ теории байесовского вывода. В результате изучения дисциплины у студента будет сформировано представление о современном состоянии дел в теории байесовского вывода. Студент получит также представление об основных методах машинного обучения, соответствующих алгоритмах вывода, вероятностных основах машинного обучения и соответствующих моделях. Изучение дисциплины будет способствовать как развитию вероятностной интуиции и разработке моделей и методов машинного обучения, так и практическому их применению.

Современные методы принятия решений

Тип дисциплины: обязательная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Современные методы анализа данных, Алгоритмы и структуры данных, Практическое программирование и анализ данных в специализированных средах

Объем з.е.: 6

Описание курса: Целями освоения дисциплины «Современные методы принятия решений» являются: формирование теоретических знаний о математических методах поиска и анализа данных для принятия и реализации решений; формирование представления о системах поддержки принятия решений. В процессе освоения дисциплины изучаются следующие темы: графические модели, основы нейронных сетей, специальные виды нейронных сетей, выбор стратегий с помощью нейросетей.

Алгоритмы и структуры данных

Тип дисциплины: обязательная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Линейная алгебра, Математический анализ, Теория вероятностей и математическая статистика.

Объем з.е.: 3

Описание курса: Целью освоения дисциплины «Алгоритмы и структуры данных» являются:

- изучение основных алгоритмов и структур данных, широко применяющихся в современной информатике.
- формирование представления об основных алгоритмах и структурах данных, алгоритмах на графах и строках, динамическом программировании, жадных алгоритмах.
- формирование представления об анализе алгоритмов, теории сложности алгоритмов и способах сравнения алгоритмов и структур данных между собой.

В ходе освоения программы изучаются следующие темы: числовые алгоритмы, рекуррентные соотношения, вычислительная сложность, алгоритмы сортировки, декомпозиция графов, пути в графах, жадные алгоритмы, динамическое программирование, структуры данных: список, массив, стек, очередь, хеш-таблица, очередь с приоритетами, алгоритмы на строках: поиск подстроки.

Практическое программирование и анализ данных в специализированных средах

Тип дисциплины: обязательная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Линейная алгебра, Математический анализ, Теория вероятностей и математическая статистика.

Объем з.е.: 6

Описание курса: Целями освоения дисциплины "Практическое программирование и анализ данных в специализированных средах" являются:

- приобретение студентами практического навыка построения и анализа алгоритмов, а также навыка самостоятельной их реализации посредством программирования на современных языках программирования (на примере Python) и в специализированных пакетах (R).
- формирование понимания технологии работы со сложными структурами программ и данных.
- формирование представления о парадигмах программирования в специализированных средах.
- развитие навыков практического программирования и анализа данных как на примере стандартных задач программирования (работа с деревьями, графами, обработка списков и массивов, символьное преобразование), так и в приложении к решению практических задач.
- формирование основы для дальнейшего применения в области математического и компьютерного моделирования сложных социально-экономических систем и процессов.

В результате изучения дисциплины у студента будет сформировано представление о истории создания языка Python, понятие о Python как динамически типизированном интерпретируемом языке программирования высокого уровня, особенностях синтаксиса Python, основах технологического обеспечения анализа данных, основах сбора, предварительной обработки и анализа данных в Python.

Математические основы анализа данных

Тип дисциплины: адаптационная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Линейная алгебра, Математический анализ, Теория вероятностей и математическая статистика.

Объем з.е.: 6

Описание курса:

Данная дисциплина является адаптационной дисциплиной программы необходима для обеспечения выравнивания уровня подготовки студентов для успешного освоения ими обязательных дисциплин. Дисциплина направлена повторение различных аспектов дискретной математики, линейной алгебры, математического анализа, теории вероятностей и математической статистики с ориентацией на их применение в анализе данных. Основные положения данной дисциплины могут быть использованы для изучения дисциплины «Современные методы анализа данных». В результате изучения дисциплины у студента будет сформировано представление о математических основах, необходимое для понимания методов машинного обучения и анализа данных, их ограничений, областей применения и принципов построения.

Bayesian Methods for Machine Learning

Type of the course: Elective

Prerequisites: The study of the discipline requires students to have knowledge and skills of confident work with a computer (an experienced user) and programming. It is assumed that students are familiar with the basics of discrete mathematics and algorithmization, have computer skills.

ECTS workload: 3

Course Overview: Bayesian methods are used in lots of fields: from game development to drug discovery. They give superpowers to many machine learning algorithms: handling missing data, extracting much more information from small datasets. Bayesian methods also allow us to estimate uncertainty in predictions, which is a really desirable feature for fields like medicine. When Bayesian methods are applied to deep learning, it turns out that they allow you to compress your models 100 folds, and automatically tune hyperparameters, saving your time and money.

In six weeks we will discuss the basics of Bayesian methods: from how to define a probabilistic model to how to make predictions from it. We will see how one can fully automate this workflow and how to speed it up using some advanced techniques.

We will also see applications of Bayesian methods to deep learning and how to generate new images with it. We will see how new drugs that cure severe diseases be found with Bayesian methods.

Вычислительная статистика

Тип дисциплины: по выбору

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Теория вероятностей и математическая статистика, Линейная алгебра, Математический анализ

Объем з.е.: 4

Описание курса: Целями освоения дисциплины «Вычислительная статистика» являются:

- изучение вычислительных и статистических методов обработки данных
- приобретение навыка в генерировании сложных вероятностных моделей,
- освоение программного обеспечения для статистической обработки данных,
- подготовка к самостоятельной исследовательской деятельности в области статистики данных.

Курс содержит следующие разделы: моделирование случайных величин и векторов; методы Монте-Карло; марковские цепи; МСМС; ресемплинг: метод складного ножа и бутстреп; улучшение и объединение оценок; ресемплинг: перестановочные тесты для проверки различных статистических гипотез; проверка гипотез с помощью бутстреп.

Advanced Econometrics

Type of the course: Elective

Prerequisites: Students' knowledge of the foundations of statistics and econometrics is a key prerequisite for the successful completion of the course.

ECTS workload: 4

Course Overview: The course is designed for first-year graduate (Master) students following the programs "Finance" and "Applied Economics and Mathematical Methods". Its main goal is to familiarize the students with advanced methods of econometric research in economics and finance. In particular, the course accentuates the problem of endogeneity and the ways to address it in the analysis of cross-sectional and panel data. The course is of applied nature: The material is presented, whenever possible, in a non-technical way, examples of empirical studies published in leading international economics and finance journals are discussed, and the lectures are supplemented by exercises in the computer lab.

The topics covered include: A review of the classical linear regression model; Causes and consequences of endogeneity; Instrumental variables methods; Key panel data techniques; Difference-in-difference estimation techniques; An overview of the matching models and regression discontinuity designs. Computer exercises using the statistical software package "Stata" are an integral part of the course, which ensures that the students get hands-on experience of analyzing real world data.

Базы данных и аналитические системы

Тип дисциплины: по выбору

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Изучение дисциплины требует от студентов знаний и навыков уверенной работы с компьютером (опытный пользователь) и программирования. Предполагается, что студенты ознакомлены с основами дискретной математики и алгоритмизации, имеют навыки работы на компьютере.

Объем з.е.: 4

Описание курса: Дисциплина посвящена изучению теоретических основ, практических методов и средств построения баз данных, а также вопросов, связанных с жизненным циклом, поддержкой и сопровождением баз данных. Рассматриваются основные понятия баз данных, способы их классификации, принципы организации структур данных и соответствующие им типы систем управления базами данных (СУБД). Подробно изучается реляционная модель данных, соответствующие этой модели СУБД, стандартный язык запросов к реляционным СУБД - SQL, методы представления сложных структур данных средствами реляционной СУБД. Рассматриваются вопросы организации коллективного доступа к данным, вводятся понятия ссылочной целостности и семантической целостности данных, транзакций, блокировок, связанные с ними проблемы и методы их решения.

Distributed processing and Big Data analysis

Type of the course: Elective

Prerequisites: Modern methods of data analysis, Mathematical foundations of data analysis, Practical programming and data analysis in specialized environments

ECTS workload: 4

Course Overview: The objectives of the discipline are:

- * familiarization of students with the basic methods of big data processing
- * formation of understanding of the internal structure, mechanics of work, the field of applicability of existing solutions;
- * students gain practical skills in analyzing large amounts of information.

In the course of studying the discipline discusses issues related to databases and databases, Sqlnosql, MapReduce model, data flows. We also study the basics of systems Hadoop, Sparkand others; algorithms for big data: clustering, downsizing, popular subject sets and associative rules; algorithms for big data: recommendation systems and Internet advertising; algorithms for big data: analysis and processing of data from social networks; the use of algorithms for processing big data in decision-making problems; architecture of big data processing systems.

Theory of economic mechanisms

Type of the course: Elective

Prerequisites: Modern methods of data analysis, Algorithms and data structures, Computational statistics.

ECTS workload: 4

Course Overview: The objectives of the discipline is:

- * developing theoretical analysis skills by solving problems and analyzing real or stylized situations and empirical market data
- * formation of knowledge about modern forms of auctions.

The course "Theory of economic mechanisms" includes the following sections: game theory: equilibrium in the Bayesian-Nash; introduction to mechanism design: definitions, principle of preference revelation; the first auction and the second price; the theorem of equivalence of profitability; efficient and optimal machinery; arrow's theorem and Gibbard-Satterthwaite; bilateral trade; williams theorem; auctions with dependent values.

Анализ социальных и экономических сетей

Тип дисциплины: по выбору

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Современные методы анализа данных, Алгоритмы и структуры данных, Математические основы анализа данных.

Объем з.е.: 4

Описание курса: Целями освоения дисциплины являются:

- знакомство студентов с подходами к анализу социальных и экономических сетей;
- формирования навыков практической работы с графами, их визуализация и обработка.

Курс предусматривает изучение следующих тем: введение в сетевой анализ; случайные графы и модели сетей; центральность и другие метрики на сетях; сообщества в сетях; влияние на сетях; визуализация; предсказание связей.

Information Retrieval and Natural Language Processing

Type of the course: Elective

Prerequisites: Modern methods of data analysis, Modern methods of decision-making

ECTS workload: 4

Course Overview: The purpose of the discipline is to familiarize students with the methods of text processing in natural language, as well as methods of processing semi-structured data and information extraction. It is assumed to get acquainted with the methods of extracting relationships, sentiment analysis, annotation and clustering of texts, as well as with existing software implementations of these methods. The course includes the following sections: introduction to natural language processing, classification and clustering of texts, information retrieval, indexing of texts for information retrieval, introduction to machine translation, introduction to information extraction, machine learning methods in the problem of information extraction, opinion extraction.

Научно-исследовательская практика

Тип дисциплины: обязательная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты): Вероятностные методы моделирования, Современные методы анализа данных.

Объем з.е.: 9

Описание курса: Целями проведения практики являются формирование навыков подготовки и проведения научных исследований, а также умения представлять полученные результаты. В результате освоения материалов по практике у студента будут сформированы навыки работы с документацией и научной литературой. Студент получит представление о процессе проведения исследований, как индивидуально, так и в коллективе. Изучение дисциплины будет способствовать повышению математической культуры студентов, даст навыки стохастического «модельного» мышления. Практика ориентирована на формирование у студентов навыков проведения самостоятельных исследований. Основной упор делается на самостоятельную работу студентов. В ходе прохождения практики студенты должны выполнить задания индивидуально.

Научно-исследовательский семинар "Машинное обучение и приложения"

Тип дисциплины: обязательная

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты):

Объем з.е.: 9

Описание курса: Целями освоения дисциплины являются:

- развитие навыков участия в научно-исследовательской деятельности;
- знакомство с современными методами машинного обучения и их практическими применениями;
- развитие навыков презентации результатов исследования и оформления презентационных материалов.

Настоящая дисциплина относится к циклу дисциплин проектной и исследовательской работы и блоку дисциплин, обеспечивающих магистерскую подготовку.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении любых дисциплин, связанных с научно-исследовательской деятельностью студентов, а так же при подготовке ВКР.

Дисциплина включает в себя следующие разделы: классические методы машинного обучения и практические прикладные задачи; глубокое обучение и его применение; технологические вопросы применения методов машинного обучения.

Научно-исследовательский семинар "Вероятностные методы моделирования"

Тип дисциплины: научно-исследовательская работа

Требования к уровню знаний студентов, необходимых для освоения дисциплины (пререквизиты):

Объем з.е.: 6

Описание курса: Целями освоения дисциплины являются:

- научить студентов азам научно-исследовательской деятельности;
- научить студентов структурировать исследование и взаимосвязывать различные разделы исследования;
- научить студентов выделять самое основное для презентации результатов исследования и оформлять презентационные материалы.

Настоящая дисциплина относится к циклу дисциплин проектной и исследовательской работы и блоку дисциплин, обеспечивающих магистерскую подготовку.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении любых дисциплин, связанных с научно-исследовательской деятельностью студентов, а также при подготовке ВКР.

Дисциплина включает в себя следующие разделы: научно-исследовательская работа студентов – виды, содержание, особенности ; выбор направления и формулировка темы исследования; постановка целей и задач; работа с источниками, цитирование, оформление ссылок и списка литературы; эмпирические/полевые/иные исследования - сбор материалов для практической части работы; методы и модели – особенности, выбор, использование, совмещение; структура работы, логика и взаимосвязь, использование иллюстративного материала, оформление; представление итогов - речь, презентация, раздаточные материалы, правила выступления; введение в проблемы дата майнинга; алгоритмы сетевого анализа; восстановление скрытых распределений пользователей в Вконтакте; медиа – войны в интернете; междисциплинарные исследования; выделение паттернов поведения из больших данных; исследование Инстаграм; анализ профилей пользователя и выявление скрытых особенностей; применение классификаторов для предсказания котировок; цели и задачи классификации медицинских данных; обзор методов Sentiment analysis; требования и структура исследовательского проекта; методы исследования; содержание и логика научной работы; обсуждение будущей письменной работы и её защиты; защита и презентация научной работы; подготовка, защита, презентация научной работы.