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

Факультет компьютерных наук

Департамент анализа данных и искусственного интеллекта

\_\_\_\_\_\_\_\_\_\_\_\_\_

**Майнор** Интеллектуальный анализ данных

**Рабочая программа дисциплины** Прикладные задачи анализа данных

для уровня подготовки – бакалавриат

Разработчики программы

Игнатов Д.И., к.т.н., доцент, dignatov@hse.ru

Черняк Е.Л., старший преподаватель, echernyak@hse.ru

Соколов Е.А., старший преподаватель, esokolov@hse.ru

Одобрена Академическим советом ОП бакалаврской программы «Прикладная математика и информатика»

 Протокол № \_\_\_\_\_\_\_\_\_\_\_\_\_\_ «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 2016 г.

Утверждена «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 2016 г.

Академический руководитель образовательной программы/руководитель подразделения (название ОП/подразделения, **предлагающей(его) майнор**)

Конушин Антон Сергеевич \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ [подпись]

Москва, 2016

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

# 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, выбравших майнор «Интеллектуальный анализ данных».

Программа разработана в соответствии с образовательным стандартом Федерального государственного автономного образовательного учреждения высшего профессионального образования «Национальный исследовательский университет «Высшая школа экономики» по направлению подготовки 01.03.02 Прикладная математика и информатика для квалификации «бакалавр».

# Цели освоения дисциплины

Целями освоения дисциплины «Прикладные задачи анализа данных» являются овладение студентами моделями и методами интеллектуального анализа данных и машинного обучения в задачах поиска информации (Information Retrieval), обработки и анализа данных, в частности, обнаружения знаний и поиска закономерностей в данных (knowledge Discovery in Databases and Data Mining), а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

# Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

* Знать основные модели и методы машинного обучения и разработки данных
* Уметь адекватно применять указанные модели и методы, а также программные средства, в которых они реализованы
* Иметь навыки (приобрести опыт) анализа реальных данных с помощью изученных методов

В результате освоения дисциплины студент осваивает следующие компетенции:

|  |  |  |  |
| --- | --- | --- | --- |
| Компетенция | Код по ФГОС/ НИУ | Дескрипторы – основные признаки освоения (показатели достижения результата) | Формы и методы обучения, способствующие формированию и развитию компетенции |
| Способен рефлексировать (оценивать и перерабатывать) освоенные научные методы и способы деятельности. | СК-М1 | Понимание места и ценности методов машинного обучения и разработки данных в современной науке и практической деятельности. | Лекции и практикумы. |
| Способен анализировать, верифицировать, оценивать полноту информации в ходе профессиональной деятельности, при необходимости восполнять и синтезировать недостающую информацию. | СК-М6 | Умеет самостоятельно находить научно-техническую литеру по тематике и критически подходить к оценке моделей и методов машинного обучения и разработки в ней рассмотренных. | Групповая работа над анализом реальных данных. |
| Способен задавать, транслировать правовые и этические нормы в профессиональной и социальной деятельности. | СЛК-М2 | Относится с уважением к международным этическим стандартам сохранения анонимности персональных данных (Data Privacy Preserving). | Работа над проектом по анализу реальных данных. |
| Способен публично представлять результаты профессиональной деятельности (в том числе с использованием информационных технологий). | ИК-М2.5 | Умеет представлять результаты проведенного исследования в виде отчета или устного рассказа с показом мультимедийной презентации. | Подготовка и представление результатов отчета по проекту. |
| Способен использовать в профессиональной деятельности знания в области естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой. | ИК-М7.1пми | Корректно применяет полученные ранее знания в таких дисциплинах как математический анализ, дифференциальные уравнение, дискретная математика и стохастическое моделирование при изучении материалов курса. | Практикумы. Работа над проектом по анализу реальных данных. |
| Способен строить и решать математические модели в соответствии с направлением подготовки и специализацией. | ИК-М7.2пми | Адекватно определяет тип задачи, строит модель и подбирает параметры методов. | Практикумы. Работа над проектом по анализу реальных данных. |
| Способен понимать и применять в исследовательской и прикладной деятельности современный математический аппарат. | ИК-М7.3пми | Корректно применяет математические модели и методы прикладной математики в анализе данных. | Работа над проектом по анализу реальных данных. |
| Способен применять в исследовательской и прикладной деятельности современные языки программирования и языки манипулирования данными, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии и т.п. | ИК-М7.5пми | Способен адекватно разрабатывать программы на одном из доступных языков программирования, использовать программные средства (в том числе самостоятельно разработанные) при сборе, предобработке и анализе данных. | Практикумы. Работа над проектом по анализу реальных данных. |

# Место дисциплины в структуре образовательной программы

Для специализаций подготовки бакалавров настоящая дисциплина является неотъемлемой дисциплиной в рамках майнора «Интеллектуальный анализ данных», который предоставляет дополнительную специализацию.

Изучение данной дисциплины базируется на следующих дисциплинах:

* Введение в программирование

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

* Необходимо владеть основами программирования на языке Python (изучается в рамках первой дисциплины курса «Введение в программирование») и знаниями математики в объеме программы средней школы.
* [Введение в машинное обучение](http://electives.hse.ru/minor_intel/machine/)
* [Интеллектуальные системы](http://electives.hse.ru/minor_intel/systems/)

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

# Тематический план учебной дисциплины

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название раздела | Всего часов  | Аудиторные часы | Самостоятельная работа |
| Лекции | Семинары | Практические занятия |
| 1 | Введение, знакомство с содержанием курса |  | 2 | 2 |  |  |
| 2 | Введение в автоматическую обработку текстов. Морфологический анализ. Методы выделения ключевых слов и словосочетаний. Синтаксический анализ. |  | 2 | 2 |  |  |
| 3 | Векторная модель документа. Дистрибутивная семантика. Методы снижения размерности.  |  | 2 | 2 |  |  |
| 4 | Задачи классификации в области автоматической обработки текстов. Классификация документов. Классификация последовательностей. |  | 2 | 2 |  |  |
| 5 | Глубинное обучение и автоматическая обработка текстов. Сверточные и рекуррентные нейронные сети.  |  | 3 | 3 |  |  |
| 6 | Анализ частых множеств признаков и ассоциативных правил (повторение и дополнение) |  | 4 | 4 |  |  |
| 7 | Анализ последовательностей (sequence mining) |  | 3 | 3 |  |  |
| 8 | Анализ формальных понятий (АФП) и его приложения |  | 2 | 2 |  |  |
| 9 | Рекомендательные системы и алгоритмы |  | 4 | 4 |  |  |
| 10 |  Бикластеризация. Мультимодальная кластеризация |  | 3 | 3 |  |  |
| 11 |  Понижение размерности (SVD, BMF, NMF) и поиск скрытых факторов в рекомендательных системах. |  | 3 | 3 |  |  |
| 12 | Спектральная кластеризация |  | 3 | 3 |  |  |
| 13 | Анализ связей (алгоритм Page Rank) |  | 3 | 3 |  |  |

# Формы контроля знаний студентов

|  |  |  |  |
| --- | --- | --- | --- |
| Тип контроля | Форма контроля | 2 год | Параметры |
| 3 | 4 |  |
| Текущий(неделя 2) | Домашнее задание | 1 |  | срок выполнения и требования указываются в тексте задания |
| Текущий(неделя 3) | Домашнее задание | 1 |  | срок выполнения и требования указываются в тексте задания |
| Текущий(неделя 4) | Домашнее задание | 1 |  | срок выполнения и требования указываются в тексте задания |
| Текущий(неделя 5) | Домашнее задание | 1 |  | срок выполнения и требования указываются в тексте задания |
| Текущий(неделя 7) | Домашнее задание |  | 1 | срок выполнения и требования указываются в тексте задания |
| Текущий(неделя 8) | Домашнее задание |  | 1 | срок выполнения и требования указываются в тексте задания |
|  | Проект по анализу реальных данных |  |  | Предполагается промежуточный контроль в конце первого модуля и итоговая защита в рамках экзамена во втором модуле. |
| Промежуточный | Техническое задание по проекту | 1 |  | В письменной форме |
| Промежуточный | Отчет по проекту |  | 1 | В письменной форме |
| Итоговый | Экзамен  |  | 1 | В форме устной защиты проекта, собеседования по результатам выполнения домашних заданий и проекта |

## Критерии оценки знаний, навыков

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

Выполнение домашнего задания оценивается в баллах по 10-ти балльной шкале. В тексте задания могут быть указаны баллы для каждого из подзаданий. Основные критерии: корректность и полнота представленного решения.

Проект по анализу реальных данных подразумевает коллективное или индивидуальное выполнение задания, связанного с применением программных средств и методов разработки данных и машинного обучения на реальных данных. Защита проекта происходит в два этапа, в первом модуле курса происходит выбор данных, обоснование методов для решения задачи и первичное исследование данных, представленных в виде технического задания (ТЗ). К защите проекта (экзамену) студент предоставляет (высылает в электронной форме) отчет по проекту (пояснительная записка, ПЗ), а в рамках устного собеседования на экзамене происходит его защита (презентация проекта, ПП).

Осуществляется следующая дистанционная поддержка при проведении контроля: задания и материалы лекций, рассылаются по электронной почте, учебная литература и материалы лекций доступны на странице курса Wiki Факультета компьютерных наук, а также в DropBox (или LMS, GitHub, SageMath и прочих облачных сервисах).

# Содержание дисциплины

Раздел 1. Введение в автоматическую обработку текстов. Морфологический анализ. Методы выделения ключевых слов и словосочетаний. Синтаксический анализ.

Краткий экскурс в историю автоматической обработки текстов (АОТ). Основные исследовательские и прикладные задачи. 7-уровневая модель языка. Основные этапы обработки текстов: графематический, морфологический и синтаксический анализ. Токенизация текста по правилам. Задачи морфологического анализа: лемматизация и определение частеречных характеристик. Задача выделения ключевых слов и словосочетаний. Меры связности. Меры контрастности. Морфологические шаблоны. Синтаксический анализ. Грамматика составляющих. Универсальные зависимости.

***Литература для освоения раздела***

* Jurafsky Daniel, Martin James H. Speech and language processing an introduction to natural language processing, computational linguistics, and speech, 2000, Главы 10, 12, 14

Раздел 2. Векторная модель документа. Дистрибутивная семантика. Методы снижения размерности.

Векторная модель документа и ее использование в задачах информационного поиска. Модель “мешок слов”. Вычисление косинусной близости между векторами. TF-IDF преобразование. Геометрическая интерпретация косинусной меры сходства. Простейший поиск по запросу. Векторная модель семантики. Дистрибутивная семантика и ее приложения. Понятие “контекста”. Снижение размерности. Сингулярное разложение. Латентно-семантический анализ (ЛСА). Использование ЛСА в задаче поиска по запросу. Использование ЛСА для визуализации дистрибутивной модели.

***Литература для освоения раздела***

* Jurafsky Daniel, Martin James H. Speech and language processing an introduction to natural language processing, computational linguistics, and speech, 2000, Главы 15, 16

Раздел 3. Задачи классификации в области автоматической обработки текстов. Классификация документов. Классификация последовательностей.

Задачи классификации, которые возникают в области АОТ: классификация документов и классификация последовательностей. Частеречный разбор как задача классификации последовательностей. Общая формулировка задачи извлечения именованных сущностей. IOB-разметка. Общие подходы к задаче классификации документов: напоминание. Метод наивного Байеса. Логистическая регрессия. Генеративные и дискриминативные методы классификации. Общие подходы к задаче классификации последовательностей. Скрытые цепи Маркова. Условные случайные поля.

***Литература для освоения раздела***

* Jurafsky Daniel, Martin James H. Speech and language processing an introduction to natural language processing, computational linguistics, and speech, 2000, Главы 6, 7, 9

Раздел 4. Глубинное обучение и автоматическая обработка текстов. Сверточные и рекуррентные нейронные сети.

Почему так сложно формулировать лингвистические признаки. Почему глубинное обучение помогает находить скрытые закономерности данных. Обзор задач автоматической обработки текстов, решаемых с помощью глубинного обучения. Современные показатели качества моделей глубинного обучения.

Модель однослойного перцептрона Маккалока-Питтса. Модель многоуровневого перцептрона. Алгоритм обратного распространения ошибки и градиентный спуск.

Классификация документов и проблема нефиксированной длины входа. Сверточная нейронная сеть. Фильтр. Макспулинг.

Вероятностные языковые модели. Рекуррентная нейронная сеть. Проблема исчезающих градиентов. Модели с длинной долгосрочной памятью.

Фреймворки глубинного обучения: Theano, Tensorflow, Keras.

***Литература для освоения раздела***

* Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016, Часть 2, главы 6-12

Раздел 5. Анализ частых множеств признаков и ассоциативных правил (повторение и дополнение)

Задача анализа потребительской корзины. Поддержка и достоверность. Частые, замкнутые и максимальные частые множества. Алгоритмы Априори и FP-growth. Импликации и ассоциативные правила как объектно-признаковые зависимости. Меры “интересности правил”. Прикладные задачи: анализ посещаемости сайтов и рекомендация контекстной рекламы. Программные средства: Orange, SPMF, ConExp.

Раздел 5. Анализ последовательностей (sequence mining).

Задача поиска частых последовательностей как развитие идеи поиска частых множеств. Форматы данных и алгоритмы (SPADE, PrefixSpan). Реализация алгоритмов в SPMF. Понятие о “контрастных” закономерностях (emerging patterns) в задачах классификации. Прикладные задачи: анализ демографических последовательностей, конкурс Recommender Systems Challenge 2015.

Раздел 6. Анализ формальных понятий (АФП) и его приложения

Базовые понятия АФП: контекст, операторы Галуа, формальное понятие, решетка понятий, импликации. Связь импликаций и ассоциативных правил. Связь частых замкнутых множеств и понятий. Диаграмма решетки понятия как средство визуализации и исследования данных. Исследование признаков (Attribute Exploration). Задача построения таксономий предметных областей. Приложения АФП в информационном поиске и анализе текстов. Поиск документов-дубликатов.

Раздел 7. Рекомендательные системы и алгоритмы

Классификация рекомендательных систем. Контентные рекомендательные системы. Коллаборативная фильтрация: сходство по пользователям (user-based) и сходство по признакам (item-based). Оценка качества рекомендательных систем (Precision, Recall, F1-мера, HitRate, MAE, RMSE, NDCG). Бимодальная кроссвалидация. Прикладная задача: рекомендация фильмов на примере данных проекта MovieLens. Альтернативные подходы оценки качества в бизнес-задачах.

Раздел 8. Бикластеризация. Мультимодальная кластеризация

Методы бикластеризация. Объектно-признаковая бикластеризация. Кластеризация по подпространству. Системы совместного пользования ресурсами и фолксономии (Folksonomy). Трикластеризация. Прикладные задачи: анализ данных экспрессии генов, социальных сетей и систем совместного пользования ресурсами. Понятие о мультимодальной кластеризации.

Раздел 9. Понижение размерности (SVD, BMF, NMF) и поиск скрытых факторов в рекомендательных системах.

Разложение матриц на основе сингулярных чисел (Singular Value Decomposition, SVD). Неотрицательная матричная факторизация (Non-negative Matrix Factorisation). Булева матричная факторизация (Boolean Matrix Factorisation). SVD-подобные модели (Funk SVD, Koren's SVD, Matrix Factorization Machines). Связь задачи рекомендаций, регрессии и классификации. Прикладная задача: рекомендация фильмов на основе матричной факторизации. Гибридные рекомендательные системы на примере рекомендации радиостанций.

9 Спектральная кластеризация

Кластеризация графов. Понятие о минимальном разрезе графа (min-cut). Преобразование Лапласа. Коэффициент Релея, вектор и число Фидлера. Иерархическая спектральная кластеризация. Прикладные задачи: поиск сообществ в соцсетях, сегментация рынка контекстной рекламы. Связь спектральной кластеризации с SVD.

10 Анализ связей (алгоритм Page Rank)

Задача ссылочного ранжирования. Алгоритмы HITS и PageRank.

# Образовательные технологии

В данной части курса первостепенными технологиями обучения являются информационно-коммуникационные технологии (ИКТ), проектные и исследовательские методы в обучении.

Необходимое для выполнения работ программное обеспечение, как правило, находится в свободном доступе и его можно загрузить в сети Интернет или скопировать из репозитория, предоставляемого курсу.

В число программных пакетов входят инструменты PyData (<http://pydata.org/downloads.html>) и другие:

* Сборка Anaconda (покрывает инструменты из списка ниже и другие)
	+ Пакеты scipy и numpy.
	+ Pandas
	+ Scikit-learn
	+ Natural Language Toolkit (NLTK)
	+ Gensim
* Keras

К инструментам с графическим интерфейсом, не требующим непосредственных навыков программирования (хотя и это возможно), относятся: Concept Explorer, SPMF, Orange, Weka и др.

Дополнительно к каждой из тем доступны слайды лекций, изложение которых адаптировано с учетом используемых программных пакетов.

В рамках этой части курса активно используется проектная форма работы.

## Методические рекомендации преподавателю

Преподавателю рекомендуется использовать демонстрацию работы изучаемых методов анализа данных с помощью предустановленных программных продуктов (PyData) и других открытых средств во время практических и лекционных занятий.

В рамках семинаров возможно решение задач, а домашние задания предлагается составлять практико-ориентированными.

Для проектов по анализу данных рекомендуется предлагать доступные источники данных, например, UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/> или <https://www.openml.org>). Для проектов по анализу текстов доступны следующие источники: коллекции текстов на русском языке – материалы конкурсов Диалог (<http://www.dialog-21.ru/evaluation/>), Открытый корпус (opencorpora.org), материалы конкурсов BSNLP (<http://bsnlp-2017.cs.helsinki.fi/shared_task.html>).

Оценочные средства для текущего контроля и аттестации студента могут включать учет работы в рамках семинарских/практических занятий.

## Тематика заданий текущего контроля

Примерные темы домашних заданий:

Домашнее задание 1. Пакет NLTK для первичной обработки текстов

Домашнее задание 2. Пакет Gensim для снижения размерности в векторной модели.

Домашнее задание 3. Пакет Keras для обучения нейронных сетей.

Домашнее задание 9. Поиск частых множеств и ассоциативных правил.

## Вопросы для оценки качества освоения дисциплины

Примерный перечень вопросов к экзамену (защите проекта):

1. Назовите две основные задачи морфологического анализа
2. По данному предложению постройте дерево составляющих
3. По данному предложению постройте дерево зависимостей
4. Что оценивает мера связности биграмм? Какие меры связности вы знаете?
5. Зачем нужно tf-idf преобразование?
6. Почему возникает необходимость снижения размерности в векторной модели?
7. Что такое word embedding?
8. Какие виды word embedding’ов вы знаете?
9. Приведите пример задачи классификации документов
10. Приведите пример задачи классификации последовательностей
11. Что общего между методом наивного Байеса и скрытыми Марковскими цепями?
12. Что общего между логистической регрессией и условными случайными полями?
13. Что такое персептрон?
14. В чем заключается метод обратного распространения ошибки?
15. Как использовать сверточные нейронные сети для классификации документов?
16. Как использовать рекуррентные нейронные сети для генерации текстов?

# Порядок формирования оценок по дисциплине

*ИО =0,5 ДЗ + 0.1 ТЗ + 0.2 ПЗ + 0.2 ПП, где*

ДЗ – домашние задания (по усмотрению семинариста оценка по домашнему заданию может быть скорректирована за счет учета работы на семинарах и практических занятиях),

ТЗ – техническое задание к проекту (индивидуальный или групповой проект, допускающий не более 3 участников),

ПЗ – пояснительная записка (письменный отчет по проекту),

ПП – презентация проекта (защита проекта).

Cпособ округления итоговой оценки по дисциплине: арифметический.

В случае особых обстоятельств студент может получить возможность пересдать низкие результаты за проектную работу.

На пересдаче студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за домашние задания.

На защите проекта студент может получить дополнительные вопросы (дополнительную практическую задачу, решить к пересдаче домашнее задание), ответ на каждый из который оценивается в 1 балл.

В диплом выставляется результирующая оценка по курсу Майнор “Интеллектуальный анализ данных”, определяемая программой первого года курса. Дисциплина "Прикладные задачи анализа данных" не дает вклад более 0,25 в результирующую оценку по курсу Майнор “Интеллектуальный анализ данных”.

# Учебно-методическое и информационное обеспечение дисциплины

## Базовые учебники

1. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014 (<http://www.dataminingbook.info/pmwiki.php/Main/BookDownload>)
2. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman: Mining of Massive Datasets, 2nd Ed. Cambridge University Press 2014 (<http://www.mmds.org>)
3. Jurafsky Daniel, Martin James H. Speech and language processing an introduction to natural language processing, computational linguistics, and speech, 2000.

(<https://web.stanford.edu/~jurafsky/slp3/>)

## Основная литература

1. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014 (<http://www.dataminingbook.info/pmwiki.php/Main/BookDownload>)
2. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman: Mining of Massive Datasets, 2nd Ed. Cambridge University Press 2014 (<http://www.mmds.org>)
3. Manning, Christopher D., and Hinrich Schütze. Foundations of statistical natural language processing. Cambridge: MIT press, 1999 (<http://nlp.stanford.edu/fsnlp/>)
4. Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016 (<http://www.deeplearningbook.org>)
5. [Francesco Ricci](http://dblp.uni-trier.de/pers/hd/r/Ricci%3AFrancesco), [Lior Rokach](http://dblp.uni-trier.de/pers/hd/r/Rokach%3ALior), [Bracha Shapira](http://dblp.uni-trier.de/pers/hd/s/Shapira%3ABracha): Recommender Systems Handbook. Springer 2015 ([http://link.springer.com/book/10.1007%2F978-1-4899-7637-6](http://link.springer.com/book/10.1007/978-1-4899-7637-6))

## Дополнительная литература

1. Òscar Celma: Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space. Springer 2010 ([http://link.springer.com/book/10.1007%2F978-3-642-13287-2](http://link.springer.com/book/10.1007/978-3-642-13287-2))
2. Boyd, Vandenberghe. Convex Optimization (<http://stanford.edu/~boyd/cvxbook/>)
3. Dekking, F.M., Kraaikamp, C., Lopuhaä, H.P., Meester, L.E., A Modern Introduction to Probability and Statistics (<http://www.ewi.tudelft.nl/index.php?id=50508> и <http://www.springer.com/gp/book/9781852338961>)
4. Boris Mirkin. Core Concepts in Data Analysis: Summarization, Correlation, Visualization. 2010 (<http://www.hse.ru/data/2010/10/14/1223126254/Mirkin_All.pdf>)

## Справочники, словари, энциклопедии

Портал [http://www.machinelearning.ru](http://www.machinelearning.ru/), энциклопедия рекомендательных систем <http://www.recsyswiki.com>.

## Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства: язык программирования Python, его библиотеки NumPy, SciPy, Pandas, Scikit-Learn, а также свободно-распространяемые библиотеки анализа данных и машинного обучения, такие как Orange, Weka, SPMF, Concept Explorer и др.

## Дистанционная поддержка дисциплины

Дистанционная поддержка может осуществляться с помощью LMS, Wiki ФКН ([http://wiki.cs.hse.ru/Майнор\_Интеллектуальный\_анализ\_данных/Прикладные\_задачи\_анализа\_данных](http://wiki.cs.hse.ru/%D0%9C%D0%B0%D0%B9%D0%BD%D0%BE%D1%80_%D0%98%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85/%D0%9F%D1%80%D0%B8%D0%BA%D0%BB%D0%B0%D0%B4%D0%BD%D1%8B%D0%B5_%D0%B7%D0%B0%D0%B4%D0%B0%D1%87%D0%B8_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85)) или хранилища слайдов и данных в DropBox папке авторов курса (возможно использование иных облачных сервисов).

# Материально-техническое обеспечение дисциплины

Используется проектор (для лекций или семинаров), слайды мультимедийных презентаций и компьютеры с предустановленным программным обеспечением и доступом в Интернет.