|  |
| --- |
| утверждаю  Руководитель Департамента  анализа данных, принятия решений и финансовых технологий  \_\_\_\_\_\_\_\_\_\_\_\_ В.И. Соловьев  «15» октября 2019 г. |
|  |

ПРИЛОЖЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

**ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ**

**И МАШИННОЕ ОБУЧЕНИЕ**

НАПРАВЛЕНИЕ ПОДГОТОВКИ: **01.03.02 «ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА»**

ПРОФИЛЬ: **АНАЛИЗ ДАННЫХ И ПРИНЯТИЕ РЕШЕНИЙ В ЭКОНОМИКЕ И ФИНАНСАХ**

ФОРМА ОБУЧЕНИЯ: **Очная**

ГОД ПРИЕМА **2017, 2018, 2019, 2020**

Год утверждения программы: 2019 год

*Одобрено департаментом анализа данных, принятия решений и финансовых технологий*

*Протокол № 3 от 15.10. 20919 г.*

**Содержание Приложения**

|  |  |
| --- | --- |
| **Наименование разделов РПД** | **стр.** |
| Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине | 2 |
| Место дисциплины в структуре образовательной программы | 4 |
| Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся (в семестре, в сессию) | 4 |
| Учебно-тематический план | 5 |
| Содержание семинаров, практических занятий | 8 |
| Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы | 10 |
| Перечень вопросов, заданий, тем для подготовки к текущему контролю | 12 |
| Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине | 13 |
| Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем | 18 |

**2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине**

**2017, 2018 год приема**

|  |  |  |  |
| --- | --- | --- | --- |
| **Код компе**  **тенции** | **Наименование**  **компетенции** | **Индикаторы достижения компетенции** | **Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/ индикаторами достижения компетенции** |
| **ПК-2** | Способность понимать, совершенствовать и применять современный математический аппарат | - | **Знать**: современный математический аппарат  **Уметь**: применять современный математический аппарат  **Владеть**: методами совершенствования и применения современного математического аппарата. |
| **ПК-7** | Способность к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения | - | **Знать:** алгоритмические и программные решения в области системного и прикладного программного обеспечения  **Уметь:** применять алгоритмические и программные решения в области системного и прикладного программного обеспечения  **Владеть:** современными технологиями разработки программных решений в области системного и прикладного программного обеспечения |
| **ПКП-4**  **(2017 г.)** | Способность понимать математические основы прикладных программ, разрабатывать вычислительные алгоритмы для решения задач в области экономики и финансов | - | **Знать:** математические основы прикладных программ,  **Уметь:** разрабатывать вычислительные алгоритмы для решения задач в области экономики и финансов  **Владеть:** технологиямиразработки вычислительных алгоритмов для решения задач в области экономики и финансов |
| **ПКП-4**  **(2018 г.)** | Способность применять математический аппарат при разработке вычислительных алгоритмов для решения задач в области экономики и финансов | - | **Знать:** математический аппарат в области экономики и финансов  **Уметь:** разрабатывать вычислительные алгоритмы  **Владеть:** навыком разработки вычислительных алгоритмов для решения задач в области экономики и финансов |
| **ПКП-5** | Способность применять методы и инструменты анализа данных при подготовке аналитического обоснования финансово-экономических решений | - | **Знать:** методы и инструменты анализа данных  **Уметь:** применять методы и инструменты анализа данных  **Владеть:** методами и инструментами анализа данных при подготовке аналитического обеспечения финансово-экономических решений |

**2019, 2020 год приема**

|  |  |  |  |
| --- | --- | --- | --- |
| **Код компе**  **тенции** | **Наименование**  **компетенции** | **Индикаторы достижения компетенции** | **Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/ индикаторами достижения компетенции** |
| ПКП-4 | Способность применять математический аппарат при разработке вычислительных алгоритмов для решения задач в области экономики и финансов | 1.Владеет математическим аппаратом, необходимым для разработки вычислительных алгоритмов.  2.Разрабатывает вычислительные алгоритмы для решения задач в области экономики и финансов. | **Знать:** математический аппарат, необходимый для МО  **Уметь**: разрабатывать вычислительные алгоритмы необходимый для МО  **Знать:** теоретические основы разработки алгоритмов МО  **Уметь**: Разрабатывать вычислительные алгоритмы для решения задач МО |
| ПКП-5 | Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений | 1.Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.  2.Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения. | **Знать:** инструментов обработки данных для решения экономических и финансовых задач в МО  **Уметь:** Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач в МО.  **Знать:** инструментов анализа данных и машинного обучения.  **Уметь:** решать прикладные задачи с использованием методов и инструментов анализа данных и машинного обучения. |

**3.** [**Место дисциплины в структуре образовательной программы**](#_Toc37342106)

Дисциплина **Технологии анализа данных и машинное обучение** относится к Модулю профиля «Анализ данных и принятие решений в экономике и финансах» направление подготовки 01.03.02 «Прикладная математика и информатика».

**4. Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся (в семестре, в сессию)**

Общая трудоемкость дисциплины составляет 4 зачетных единиц.

Вид промежуточной аттестации: экзамен

Вид текущего контроля – домашнее творческое задание, курсовая работа

*Очная форма обучения, 2017 г.*

|  |  |  |
| --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Семестр 6**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **4/144** | **144** |
| ***Контактная работа - Аудиторные занятия*** | ***54*** | ***54*** |
| *Лекции* | *18* | *18* |
| *Семинары, практические занятия* | *36* | *36* |
| ***Самостоятельная работа*** | ***90*** | ***90*** |
| Вид текущего контроля | Домашнее творческое задание, курсовая работа (20 час.) | Домашнее творческое задание, курсовая работа (20 час.) |
| Вид промежуточной аттестации | Экзамен | Экзамен |

Общая трудоемкость дисциплины составляет 8 зачетных единиц.

Вид промежуточной аттестации: экзамен (5, 6 семестр)

Вид текущего контроля: курсовая работа, домашнее творческое задание, расчетно-аналитическая работа

*Очная форма обучения, 2018 г.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Семестр 5**  **(в часах)** | **Семестр 6**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **8/288** | **144** | **144** |
| ***Контактная работа –***  ***Аудиторные занятия*** | ***100*** | ***50*** | ***50*** |
| *Лекции* | *32* | *16* | 16 |
| *Семинары, практические занятия* | *68* | *34* | 34 |
| ***Самостоятельная работа*** | ***188*** | ***94*** | ***94*** |
| Вид текущего контроля | Курсовая работа, домашнее творческое задание, расчетно-аналитическая работа | Курсовая работа (20 час.),  домашнее творческое задание | Расчетно-аналитическая работа |
| Вид промежуточной аттестации | Экзамен | Экзамен | Экзамен |

Общая трудоемкость дисциплины составляет 8 зачетных единиц.

Вид промежуточной аттестации: экзамен

Вид текущего контроля – домашнее творческое задание, курсовая работа

*Очная форма обучения, 2019/2020 г. г.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Семестр 6/5**  **(в часах)** | **Семестр 7/6**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **8/288** | **108/144** | **180/144** |
| ***Контактная работа –***  ***Аудиторные занятия*** | ***102/100*** | ***34/50*** | ***68/50*** |
| *Лекции* | *32/32* | *16/16* | 16/16 |
| *Семинары, практические занятия* | *70/68* | *18/34* | 52/34 |
| ***Самостоятельная работа*** | ***186/188*** | ***74/94*** | ***112/94*** |
| Вид текущего контроля | Домашнее творческое задание, курсовая работа | Домашнее творческое задание | Курсовая работа  (20 час.) |
| Вид промежуточной аттестации | Экзамен | Экзамен | Экзамен |

**5.2.** **Учебно – тематический план**

*Очная форма обучения, 2017 г.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Наименование тем**  **(разделов)**  **дисциплины** | **Трудоёмкость в часах** | | | | | | **Формы текущего контроля успеваемости** |
| **Всего** | **Аудиторная работа** | | | | **Самостоятельная работа** |
| Общая, в т.ч.: | Лекции | Семинары, практические занятия | Занятия в интерактивных формах |
| 1. | Информационные технологии анализа данных | 18 | 6 | 2 | 4 | 2 | 12 | Устный опрос, проверка практических заданий |
| 2. | Введение в машинное обучение и знакомство с технологиями машинного обучения | 16 | 4 | 2 | 2 | 2 | 12 | Устный опрос, проверка практических заданий |
| 3. | Методы обучения с учителем | 14 | 4 | 2 | 2 | 2 | 10 | Устный опрос, проверка практических заданий |
| 4. | Методы восстановления регрессии | 14 | 6 | 2 | 4 | 2 | 8 | Устный опрос, проверка практических заданий |
| 5. | Обучение без учителя, кластеризация | 12 | 4 | 2 | 2 | 2 | 8 | Устный опрос, проверка практических заданий |
| 6. | Коллаборативная фильтрация, отбор признаков | 14 | 6 | 2 | 4 | 2 | 8 | Устный опрос, проверка практических заданий |
| 7. | Генетические алгоритмы | 14 | 6 | - | 6 | 2 | 8 | Устный опрос, проверка практических заданий |
| 8. | Анализ сетевых структуру | 14 | 6 | 2 | 4 | 2 | 8 | Устный опрос, проверка практических заданий |
| 9. | Анализ текстов на естественном языке | 14 | 6 | 2 | 4 | 2 | 8 | Устный опрос, проверка практических заданий |
| 10. | Визуализация результатов анализа данных | 14 | 6 | 2 | 4 | 2 | 8 | Устный опрос, проверка практических заданий |
|  | В целом по дисциплине | 144 | 54 | 18 | 36 | 20 | 90 | Домашнее творческое задание, курсовая работа |
|  | Итого в % |  |  |  |  | 37% |  |  |

*Очная форма обучения, 2018, 2019/2020 г.г.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Наименование тем**  **(разделов)**  **дисциплины** | **Трудоёмкость в часах** | | | | | | **Формы текущего контроля успеваемости** |
| **Всего** | **Аудиторная работа** | | | | **Самостоятельная работа** |
| Общая, в т.ч.: | Лекции | Семинары, практические занятия | Занятия в интерактивных формах |
| 1. | Информационные технологии анализа данных | 30/30 | 12/12 | 4/4 | 8/8 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 2. | Введение в машинное обучение и знакомство с технологиями машинного обучения | 26/28 | 8/10 | ~~4~~ | 4/6 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 3. | Методы обучения с учителем | 28/28 | 10/10 | 2 | 8/8 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 4. | Методы восстановления регрессии | 28/28 | 10/10 | 2 | 8/8 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 5. | Обучение без учителя, кластеризация | 36/34 | 10/10 | 2 | 8/8 | 2 | 26/24 | Устный опрос, проверка практических заданий |
| 6. | Коллаборативная фильтрация, отбор признаков | 30/30 | 12/12 | 6 | 6/6 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 7. | Генетические алгоритмы | 28/28 | 10/10 | 2 | 8/8 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 8. | Анализ сетевых структуру | 26/26 | 8/8 | ~~4~~ | 4/4 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 9. | Анализ текстов на естественном языке | 28/28 | 10/10 | 2 | 8/8 | 2 | 18/18 | Устный опрос, проверка практических заданий |
| 10. | Визуализация результатов анализа данных | 28/28 | 10/10 | 4 | 6/6 | 2 | 18/18 | Устный опрос, проверка практических заданий |
|  | В целом по дисциплине | 288 | 100/102 | 32 | 68/70 | 20 | 188/186 | Курсовая работа, домашнее творческое задание, расчетно-аналитическая работа - для 2018 г. |
|  | Итого в % |  |  |  |  | 20% |  |  |

**5.3. Содержание семинаров, практических занятий**

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)** | **Формы проведения занятий** |
| 1.Информационные технологии анализа данных | Входной контроль. Изучение технологического стека анализа данных, построенного на базе языка программирования Python.  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 2.Введение в машинное обучение и знакомство с технологиями машинного обучения | Решение задач машинного обучения с использованием библиотеки Scikit-Learn. Оценка качества результатов: тестирование, разбиение набора данных для обучения и тестирования, кроссвалидация. Расчет метрик качества результатов.  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 3.Методы обучения с учителем | Решение задач машинного обучения с использованием метрических методов классификации при помощи библиотеки Scikit-Learn.  Решение задач машинного обучения с использованием решающих списков и деревьев при помощи библиотеки Scikit-Learn.  Решение задач машинного обучения с использованием байесовских классификаторов при помощи библиотеки Scikit-Learn.  Решение задач машинного обучения с использованием метода опорных векторов при помощи библиотеки Scikit-Learn.  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 4.Методы восстановления регрессии | Решение задач восстановления регрессии при помощи библиотеки Scikit-Learn.  Решение задач машинного обучения с использованием логистической регрессии при помощи библиотеки Scikit-Learn.  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 5.Обучение без учителя, кластеризация | Решение задач кластеризации при помощи библиотеки Scikit-Learn (алгоритм k средних (k-means), другие методы кластеризации).  Решение задач снижения размерности при помощи библиотеки Scikit-Learn (метод главных компонент).  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 6.Коллаборативная фильтрация, отбор признаков | Решение задач коллаборативной фильтрации на языке Python (корреляционные методы, латентные методы).  Реализация скользящего контроля на Python. Решение задачи отбора признаков на Python.  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 7.Генетические алгоритмы | Выполнение аналитических задач с использованием технологий генетических алгоритмов.  *Рекомендуемые источники: 8. 1-6* | Индивидуальное выполнение заданий, групповой разбор результатов выполнения заданий (33% времени на интерактивные технологии) |
| 8.Анализ сетевых структуру | Решение задач создания (в т.ч. загрузки из внешних источников и генерации) сетевых структур. Извлечение данных, выполнение запросов к графовой СУБД. Определение свойств сети, визуализация сети.  *Рекомендуемые источники: 8. 1-6* | Выполнение практических заданий на компьютере, работа в группах (33% времени на интерактивные технологии). |
| 9.Анализ текстов на естественном языке | Токенезация, нормализация текста. Векторное представление текста, решение задач с помощью алгоритма TF-IDF.  *Рекомендуемые источники: 8. 1-6* | Выполнение практических заданий на компьютере, работа в группах (33% времени на интерактивные технологии). |
| 10.Визуализация результатов анализа данных | Выполнение заданий по визуализации результатов анализа. Тематические дискуссии на темы: какие методы визуализации помогают улучшить понимание выявленных закономерностей; можно ли считать саму визуализацию инструментом анализа данных  *Рекомендуемые источники: 8. 1-6* | Выполнение практических заданий на компьютере, работа в группах (33% времени на интерактивные технологии). |

**6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы**

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов, отводимых на самостоятельное освоение** | **Формы внеаудиторной самостоятельной работы** |
| 1.Информационные технологии анализа данных | Знакомство с интерактивной оболочкой IPython notebook. Изучение принципов работы и применения для решения задач анализа данных и машинного обучения.  Знакомство с библиотеками numpy и pandas и решением базовых задач подготовительных операций для выполнения анализа данных с помощью этих библиотек. | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |
| 2.Введение в машинное обучение и знакомство с технологиями машинного обучения | Расчет метрик качества Precision, Recal, F1, RUC. Критерии качества классификации: чувствительность и специфичность, ROC-кривая и AUC, точность и полнота, AUC-PR. | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |
| 3. Методы обучения с учителем | Решение задач машинного обучения с использованием метрических методов классификации при помощи библиотеки Scikit-Learn (сравнение решения задач с различными параметрами и видами метрических алгоритмов).  Решение задач машинного обучения с использованием решающих списков и деревьев при помощи библиотеки Scikit-Learn (сравнение решения задач с различными параметрами и видами алгоритмов решающих списков и деревьев).  Решение задач машинного обучения с использованием байесовских классификаторов при помощи библиотеки Scikit-Learn (сравнение решения задач с различными параметрами и видами байесовских алгоритмов).  Решение задач машинного обучения с использованием метода опорных векторов при помощи библиотеки Scikit-Learn (сравнение решения задач с различными параметрами и видами алгоритмов метода опорных векторов). | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |
| 4.Методы восстановления регрессии | Решение задач восстановления регрессии при помощи библиотеки Scikit-Learn с штрафами L1 и L2, Elastic Net.  Решение задач машинного обучения с использованием логистической регрессии при помощи библиотеки Scikit-Learn. | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |
| 5.Обучение без учителя, кластеризация | Решение задач кластеризации при помощи библиотеки Scikit-Learn (использование методов, альтернативных алгоритму k средних).  Решение задач снижения размерности при помощи библиотеки Scikit-Learn (использование методов, альтернативных методу главных компонент). | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |
| 6.Коллаборативная фильтрация, отбор признаков | Решение задач коллаборативной фильтрации на языке Python корреляционными методами с различными метриками.  Тестирование альтернативных алгоритмов решения задачи отбора признаков. | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |
| 7.Генетические алгоритмы | Генетические алгоритмы в прикладных задачах. Преимущества и ограничения технологий генетических алгоритмов. | Работа с литературой, работа с электронными источниками |
| 8.Анализ сетевых структур | Идентификация сетей, имеющих структуру тесного мира  Визуализация сетей имеющих привязку узлов на плоскости | Работа с литературой, работа с электронными источниками |
| 9.Анализ текстов на естественном языке | Классификация тематического корпуса текстов | Работа с литературой, работа с электронными источниками |
| 10.Визуализация результатов анализа данных | Современные методы визуализации аналитической информации  Краткое описание данных на основе визуализации | Работа с литературой, работа с электронными источниками, разработка алгоритмов и программ |

**6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю**

***Примерная тематика курсовой работы***

1. Визуализация результатов анализа данных для набора данных биржевых торгов.
2. Визуализация результатов анализа данных для набора данных о корзинах покупок.
3. Визуализация результатов анализа данных для набора данных о продажах недвижимости.
4. Решение одной из задач анализа текстов на естественном языке для корпуса текстов на русском языке.
5. Решение одной из задач анализа текстов на естественном языке для корпуса текстов на английском языке.
6. Коллаборативная фильтрация для набора данных о покупках в интенет-магазине.
7. Коллаборативная фильтрация для набора данных об использовании хэш-тегов.
8. Решение задачи кластеризации для набора данных о покупках автомобилей.
9. Решение задачи кластеризации для набора данных о демографических показателях.
10. Решение задачи кластеризации для набора данных о продажах недвижимости.

**Примеры заданий расчетно-аналитической работы**

1. Визуализируйте результаты кластеризации многомерных данных для задачи классификации заемщика к категории «плохих заемщиков».
2. Методами анализа текстов на естественном языке проведите кластеризацию корпуса текстов из соцсетей для классификации заемщика к категории «плохих заемщиков».
3. Решите задачу коллаборативной фильтрации для заданного набора данных для задачи классификации заемщика к категории «плохих заемщиков».
4. Решите задачу отбора признаков для задачи поиска наиболее перспективного биржевого актива.
5. Визуализируйте результаты кластеризации многомерных данных для задачи поиска наиболее перспективного биржевого актива.
6. Методами анализа текстов на естественном языке проведите кластеризацию корпуса новостных текстов для задачи поиска наиболее перспективного биржевого актива.
7. Решите задачу коллаборативной фильтрации для заданного набора данных для задачи поиска наиболее перспективного биржевого актива.
8. Решите задачу отбора признаков для задачи для задачи поиска наиболее стабильного биржевого актива.
9. Визуализируйте результаты кластеризации многомерных данных для задачи поиска наиболее стабильного биржевого актива.
10. Методами анализа текстов на естественном языке проведите кластеризацию корпуса новостных текстов для задачи поиска наиболее стабильного биржевого актива.

**Примерная тематика домашнего творческого задания**

1. Визуализация результатов анализа данных.
2. Анализ текстов на естественном языке.
3. Коллаборативная фильтрация, отбор признаков.
4. Методы обучения без учителя.
5. Кластеризация.
6. Методы обучения с учителем.
7. Какие виды активационных функций используются в искусственной нейронной сети типа RBF (сеть с радиальной базисной функцией)?
8. С помощью какого представления карты Кохонена можно понять сколько нейронов не стали победителями ни для одного образца?
9. Какие связи присутствуют в искусственной нейронной сети типа MLP (многослойный персептрон)?

**7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2 «Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

**Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний**

**2019, 2020 год приема**

|  |  |
| --- | --- |
| **Наименование компетенций** | **Примеры заданий для оценки индикаторов достижения компетенций** |
| ПКП-4 Способность применять математический аппарат при разработке вычислительных алгоритмов для решения задач в области экономики и финансов | **1.Владеет математическим аппаратом, необходимым для разработки вычислительных алгоритмов.**  **Задание 1.**  Предложите метрику оценки качество модели коллаборативной фильтрации.  **2.Разрабатывает вычислительные алгоритмы для решения задач в области экономики и финансов.**  **Задание 2.**  Реализуйте модель для решения задачи коллаборативной фильтрации. |
| ПКП-5 Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений | **1.Осуществляет обоснованный выбор методов и инструментов обработки данных для решения экономических и финансовых задач.**  **Задание 1.**  Предложите технологический стек и архитектуру для решения задачи коллаборативной фильтрации с производительностью не менее 5000 t/s.  **2.Владеет навыками решения прикладных задач с использованием методов и инструментов анализа данных и машинного обучения.**  **Задание 2.**  Реализуйте решение задачи коллаборативной фильтрации для типичного интернет-магазина. |

**2017, 2018 год приема**

|  |  |
| --- | --- |
| **Наименование компетенций** | **Примеры заданий для оценки сформированности компетенций** |
| ПК-2 Способность понимать, совершенствовать и применять современный математический аппарат | Предложите метрику оценки качество модели коллаборативной фильтрации. |
| ПК-7 Способность к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения | Реализуйте решение задачи коллаборативной фильтрации для типичного интернет-магазина. |
| ПКП-4 (2017 г.) Способность понимать математические основы прикладных программ, разрабатывать вычислительные алгоритмы для решения задач в области экономики и финансов | Предложите метрику оценки качество модели коллаборативной фильтрации. |
| ПКП-4 (2018 г.) Способность применять математический аппарат при разработке вычислительных алгоритмов для решения задач в области экономики и финансов | Реализуйте алгоритм k ближайших соседей для классификации заемщика к категории «плохих заемщиков». |
| ПКП-5 Способность применять методы и инструменты анализа данных при подготовке аналитического обоснования финансово-экономических решений | Предложите технологический стек и архитектуру для решения задачи коллаборативной фильтрации с производительностью не менее 5000 t/s. |

**Примеры практико-ориентированных (ситуационных) заданий**

1. Проведите анализ новостного потока CNN (включая текстовые и видео сообщения) и провести статистический анализ влияния потока на биржевые показатели ключевых мировых рынков. Предложить подход к оценке качества и тестированию модели. Привести методы визуализации входных данных и результатов модели.
2. Предложите технологический стек и архитектуру для решения задачи коллаборативной фильтрации с производительностью не менее 5000 t/s. Предложить подход к оценке качества и тестированию модели. Привести методы визуализации входных данных и результатов модели.
3. Предложить пример матричного разложения для частично определённой матрицы 5 на 4, содержащей оценки по пятибалльной шкале. Показать, как с помощью разложения можно определить оценку для неопределенной ячейки матрицы. Дать содержательную интерпретацию задачи. Предложить подход к оценке качества и тестированию модели. Привести методы визуализации входных данных и результатов модели.
4. Решить задачу коллаборативной фильтрации на предложенных исходные данных (в т.ч. их матричном представлении). Корреляционные методы user-based, item-based. Различные меры сходства субъектов и объектов. Предложить подход к оценке качества и тестированию модели. Привести методы визуализации входных данных и результатов модели.
5. Для задачи классификации красных и синих объектов, предложенных на рисунке. Предложить графический вид трех моделей классификации: недообученной, перобученной и нормально обученной. Обосновать характеристики данные моделям. Предложить подход к оценке качества и тестированию модели. Привести методы визуализации входных данных и результатов модели.

**Примерные вопросы для подготовки к экзаменам**

1. Понятие машинного обучения. Оценка качества результатов машинного обучения.
2. Виды шкал признаков. Подготовка и очистка данных для машинного обучения.
3. Понятия переобучения и недообучения. Диагностика переобучения и подходы к борьбе с ним.
4. Задача классификации и задача регрессии. Постановка, обзор подходов к решению.
5. Метод логистической регрессии.
6. Метод опорных векторов.
7. Метод линейной регрессии.
8. Метод градиентного спуска.
9. Метод случайного леса и метод решающих деревьев.
10. Задача кластеризации. Постановка и формализация.
11. Задача понижения размерности. Постановка и применение.
12. Перекрестная проверка (кросс-валидация). Описание метода и применение.
13. Задача обнаружения аномалий.
14. Закон Ципфа и стоп-слова в тексте
15. Языковые модели, n-граммы, вероятностная модель с допущением Маркова
16. Расстояние Левенштейна, алгоритм Вагнера-Фишера
17. Модель перцептрона. Проблема классификации для однослойной ИНС.
18. Механизм обратного распространения ошибки.
19. Метод стохастического градиентного спуска.
20. История развития методов построения ИНС. Принципы глубокого обучения.
21. Проблема переобучения ИНС и методы регуляризации ИНС.
22. Нормализация по мини-батчам.
23. Усовершенствованные методы градиентного спуска.
24. Kеras: общие принципы, технологический стек, последовательный и функциональный стиль описания моделей.
25. ТеnsorFlow: описание архитектуры вычислений.
26. Принцип работы сверточных ИНС.
27. Преобразование размерности данных при использовании сверточных сетей и операций пулинга.
28. Типичная архитектура сети распознавания изображений с помощью сверточных ИНС.
29. Организация глубокого обучения на небольших наборах изображений.
30. Специализированные базы данных фиксирующие лексические и семантические отношения между словами.
31. Дистрибутивная семантика, матрица совместной встречаемости и меры близости векторов.
32. Представление слов в виде векторов малой размерности с помощью сингулярного разложения.
33. Принцип работы Word2Vec, описание алгоритма CBOW
34. Принцип работы Word2Vec, описание алгоритма Skip-gram

**Примеры экзаменационных билетов**

**Билет № 1**

На основе предложенного массива данных о ценах на 1, 2 и 3-х комнатные квартиры в разных округах г. Москвы постройте необходимые визуальные представления, и ответьте на следующие вопросы:

1. Как различаются между собой округа по следующим параметрам цен на квартиры: средняя, медиана, размах двух центральных квартилей, отличие средней цены первого и последнего дециля совокупности?

2. Как различается количественное предложение квартир с разным числом комнат по округам?

**Билет № 2**

На основе предложенного массива данных о продажах в супермаркете (покупательские корзины) выполните поиск шаблонов типа «ассоциация». На основе полученных данных выявите наиболее значимые с точки зрения возможности практического применения правила (5-7 правил). Обоснуйте ваш выбор на основе значений поддержки и достоверности.

**11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем**

**11. 1. Комплект лицензионного программного обеспечения:**

1. Windows, Microsoft Office.

2. Антивирус ESET Endpoint Security

**11.2. Современные профессиональные базы данных и информационные справочные системы**

1. Информационно-правовая система «Гарант»

2. Информационно-правовая система «Консультант Плюс»

3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

4.Система комплексного раскрытия информации «СКРИН» -http://www.skrin.ru/

**11.3. Сертифицированные программные и аппаратные средства защиты информации**

- не используются