

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ  
федеральное государственное автономное образовательное учреждение высшего  
образования  
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 63

УТВЕРЖДАЮ

Руководитель направления

к. ф. н., доц.

(должность, уч. степень, звание)

М.А. Чиханова

(инициалы, фамилия)



(подпись)

«23» июня 2021 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Машинное обучение»

(Наименование дисциплины)


|   |  |
|---|--|
| Код направления подготовки/<br>специальности          | 45.03.02                               |
| Наименование направления<br>подготовки/ специальности | Лингвистика                            |
| Наименование<br>направленности                        | Теоретическая и прикладная лингвистика |
| Форма обучения  | очная                                  |

Санкт-Петербург– 2021

Лист согласования рабочей программы дисциплины

Программу составил (а)

зав. каф., доц., к. филол. н.  
(должность, уч. степень,  
звание)

 25.05.2021 г.  
(подпись, дата)


М. А. Чиханова  
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 63

«25» мая 2021 г., протокол № 8

Заведующий кафедрой № 63


к. ф. н., доц.  
(уч. степень, звание)

 25.05.2021 г.  
(подпись, дата)

М.А. Чиханова  
(инициалы, фамилия)

Ответственный за ОП ВО 45.04.02(01)


доц., к. ф. н.  
(должность, уч. степень,  
звание)

 25.05.2021 г.  
(подпись, дата)

Е. Ю. Дубинина  
(инициалы, фамилия)

Заместитель декана факультета №6 по методической работе

доц., к.п.н., доц.  
(должность, уч. степень,  
звание)

 25.05.2021 г.  
(подпись, дата)

И. М. Евдокимов  
(инициалы, фамилия)

## Аннотация

Дисциплина «Машинное обучение» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/ специальности 45.03.02 «Лингвистика» направленности «Теоретическая и прикладная лингвистика». Дисциплина реализуется кафедрой «№63».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

ПК-1 «Способность работать с основными информационно-поисковыми и экспертными системами, системами представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации и автоматизированного перевода, автоматизированными системами идентификации и верификации личности»

ПК-5 «Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности»

ПК-6 «Владение методами формального и когнитивного моделирования естественного языка и методами создания метаязыков»

ПК-7 «Владение основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов»

ПК-8 «Способность решать основные типы задач в области лингвистического обеспечения информационных и других прикладных систем»

Содержание дисциплины охватывает круг вопросов, связанных с изучением некоторых алгоритмов обучения с учителем и без с использованием соответствующих функций библиотеки компьютерного зрения OpenCV и IPP. В курсе рассматриваются основные задачи обучения по прецедентам: классификация, кластеризация, регрессия, понижение размерности. Изучаются методы их решения, как классические, так и новые, созданные за последние 10–15 лет. Упор делается на глубокое понимание математических основ, взаимосвязей, достоинств и ограничений рассматриваемых методов. Отдельные теоремы приводятся с доказательствами.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия, самостоятельная работа студента, курсовое проектирование.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц, 180 часов.

Язык обучения по дисциплине русский.

## 1. Перечень планируемых результатов обучения по дисциплине

### 1.1. Цели преподавания дисциплины

Цель данной дисциплины – изучить некоторые алгоритмы обучения с учителем и без с использованием соответствующих функций библиотеки компьютерного зрения OpenCV и IPP. В курсе рассматриваются основные задачи обучения по прецедентам: классификация, кластеризация, регрессия, понижение размерности. Изучаются методы их решения, как классические, так и новые, созданные за последние 10–15 лет. Упор делается на глубокое понимание математических основ, взаимосвязей, достоинств и ограничений.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

| Категория (группа) компетенции | Код и наименование компетенции   | Код и наименование индикатора достижения компетенции  |
|--------------------------------|--|---|
| Профессиональные компетенции   | ПК-1 Способность работать с основными информационно-поисковыми и экспертными системами, системами представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации и автоматизированного перевода, автоматизированными системами идентификации и верификации личности | ПК-1. 3.1 знать основные информационно-поисковые и экспертные системы<br>ПК-1. В.1 владеть навыками работы с системами представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, автоматизированными системами идентификации и верификации личности |
| Профессиональные компетенции   | ПК-5 Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической  | ПК-5. 3.1 знать основные принципы организации информационно-поисковых систем с применением информационно-лингвистических технологий<br>ПК-5. У.1 уметь практически реализовывать основные алгоритмы информационного поиска  |

|                              |  |  |
|------------------------------|--|--|
|                              | культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности   |  |
| Профессиональные компетенции | ПК-6 Владение методами формального и когнитивного моделирования естественного языка и методами создания метаязыков   | ПК-6. 3.1 знать существующие методы когнитивного и формального моделирования естественного языка, системы обработки естественного языка и машинного перевода             |
| Профессиональные компетенции | ПК-7 Владение основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов | ПК-7. У.1 уметь с учетом элементов программирования и автоматической обработки лингвистических корпусов обрабатывать лингвистическую информацию                          |
| Профессиональные компетенции | ПК-8 Способность решать основные типы задач в области лингвистического обеспечения информационных и других прикладных систем   | ПК-8. 3.1 знать терминологию, алгоритмы и схемы, приемы и законы создания и чтения чертежей и документации по аппаратным и программным компонентам информационных систем |

## 2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

- Математика. Теория вероятностей и мат. статистика.
- Информационные технологии в лингвистике.

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и могут использоваться при изучении других дисциплин:

- Обработка текстовых массивов.
- Информационные языки.
- Автоматизированный перевод.
- Производственная преддипломная практика.

## 3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

| Вид учебной работы  | Всего  | Трудоемкость по семестрам |
|---|--------|---------------------------|
|   |        | №6                        |
| 1   | 2      | 3                         |
| <b>Общая трудоемкость дисциплины, ЗЕ/ (час)</b>   | 5/ 180 | 5/ 180                    |
| <b>Из них часов практической подготовки</b>   | 85     | 85                        |
| <b>Аудиторные занятия, всего час.</b>   | 119    | 119                       |
| в том числе:  |        |                           |
| лекции (Л), (час)   | 34     | 34                        |
| практические/семинарские занятия (ПЗ), (час)  | 68     | 68                        |
| лабораторные работы (ЛР), (час)   |        |                           |
| курсовой проект (работа) (КП, КР), (час)  | 17     | 17                        |
| экзамен, (час)  | 36     | 36                        |
| <b>Самостоятельная работа, всего (час)</b>  | 25     | 25                        |
| <b>Вид промежуточной аттестации:</b> зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**) | Экз.   | Экз.                      |

Примечание: \*\* кандидатский экзамен

#### 4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

| Разделы, темы дисциплины  | Лекции (час) | ПЗ (СЗ) (час) | ЛР (час) | КП (час) | СРС (час) |
|---|--------------|---------------|----------|----------|-----------|
| Семестр 6   |              |               |          |          |           |
| Раздел 1. Введение в машинное обучение  | 2            | 4             |          |          | 7         |
| Раздел 2. Задача обучения с учителем  | 20           | 32            |          |          | 9         |
| 2.1. Постановка задачи обучения с учителем  |              |               |          |          |           |
| 2.2. Метод k ближайших соседей  |              |               |          |          |           |
| 2.3. Машина опорных векторов  |              |               |          |          |           |
| 2.4. Деревья решений  |              |               |          |          |           |
| 2.5. Случайный лес  |              |               |          |          |           |
| 2.6. Градиентный бустинг деревьев решений   |              |               |          |          |           |
| Раздел 3. Обзор возможностей библиотеки OpenCV и IPP для решения задач обучения без учителя | 12           | 32            |          |          | 9         |
| 3.1. Кластеризация  |              |               |          |          |           |
| 3.2. Метод центров тяжести  |              |               |          |          |           |
| 3.3. Метод медиан   |              |               |          |          |           |
| Выполнение курсовой работы  |              |               |          | 17       |           |
| Итого в семестре:   | 34           | 68            |          | 17       | 25        |
| Итого   | 34           | 68            |          | 17       | 25        |
|   |              |               |          |          |           |

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

#### 4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

| Номер раздела | Название и содержание разделов и тем лекционных занятий   |
|---------------|---|
| 1             | <p><b>Введение в машинное обучение</b><br/>Область исследования, методы изучения. История вопроса. Состояние дисциплины на современном этапе. Межпредметная связь. Основные задачи индуктивного и дедуктивного обучения.</p>  |
| 2             | <p><b>Задача обучения с учителем</b><br/> <b>2.1. Постановка задачи обучения с учителем</b><br/>           Задача обучения с учителем заключается в восстановлении зависимости между входом и выходом по имеющейся обучающей выборке, т. е. необходимо построить функцию (решающее правило) <math>f: \mathcal{X} \rightarrow \mathcal{Y}</math>, по новым объектам <math>x \in \mathcal{X}</math> предсказывающую ответу <math>y = f(x) \in \mathcal{Y}</math>:<br/> <math>y = f(x) \approx f^*(x)</math>.</p> <p>Алгоритмы обучения, компоненты, свойства (атрибуты) объекта. Признаки объекта: количественным или вещественные, номинальные или категориальные, бинарные, порядковые. Пространство признаков. Задача классификации.<br/>           Задачи обучения: ранжирование, прогнозирование и др. Проблема эффективности вычислимости.</p> <p><b>2.2. Метод k ближайших соседей</b><br/>           Метод <math>k</math> ближайших соседей (k nearest-neighbor, k-NN) как наиболее простой и универсальный метод, используемые как для решения задач классификации, так и для восстановления регрессии. Классификация нового объекта посредством отнесения его к классу, являющемуся преобладающим среди <math>k</math> ближайших (в пространстве признаков) объектов из обучающей выборки. Учет весов и расстояния до объекта при решении задач с несбалансированными данными. Использование евклидова расстояния для определения ближайших соседей. Использование расстояния Хэмминга в случае качественных переменных.<br/>           Проблема повышения точности модели обучения. Число «ближайших соседей» как один из основных параметров обобщающей способности алгоритма.</p> <p><b>2.3. Машина опорных векторов</b><br/>           Один из самых популярных методов машинного обучения – машина опорных векторов (SVM – Support Vector Machine) – как развитие идей, предложенных в 1960–1970 гг. В. Н. Вапником и А. Я. Червоненкисом.<br/>           Математическая постановка задачи двух линейно разделимых классов.<br/>           Совершенствование метода делимости графов.<br/>           Использование метода опорных векторов для задачи классификации. Стратегии "Каждый против каждого" и "Один против всех".</p> <p><b>2.4. Деревья решений</b><br/>           Деревья решений как универсальный алгоритм обучения. Их достоинства, эффективные алгоритмы для их настройки. Рекурсивное построения дерева посредством «жадной» процедуры.<br/>           Глубина дерева решений. Разбиение и процедура отсечения разбиений.</p> <p><b>2.5. Случайный лес</b></p> |

|   |  |
|---|--|
|   | <p>Использовании композиции "слабых" решающих правил как один из общих подходов в машинном обучении. Построение итогового правила посредством взвешенного голосования ансамбля базовых правил. Использование двух идей для построения базовых правил и вычисления весов:</p> <ul style="list-style-type: none"> <li>• Баггинг (bagging – bootstrap aggregation): обучение базовых правил происходит на различных случайных подвыборках данных или/и на различных случайных частях признакового описания; при этом базовые правила строятся независимо друг от друга.</li> <li>• Бустинг (boosting): каждое следующее базовое правило строится с использованием информации об ошибках предыдущих правил, а именно, веса объектов обучающей выборки подстраиваются таким образом, чтобы новое правило точнее работало на тех объектах, на которых предыдущие правила чаще ошибались.</li> </ul> <p>Модификации метода случайного леса.</p> <p><b>2.6. Градиентный бустинг деревьев решений</b></p> <p>Градиентный бустинг деревьев решений универсальный алгоритм машинного обучения, основанный на использовании ансамбля деревьев решений, его отличия от случайного леса.</p> <p>Функция Хьюбера и квадратичный штраф.</p> <p>Другой популярный метод, использующий идею бустинга, - алгоритм AdaBoost и его модификации.</p> |
| 3 | <p><b>Обзор возможностей библиотеки OpenCV и IPP для решения задач обучения без учителя</b></p> <p><b>3.1. Кластеризация</b></p> <p>Задачи кластеризации, понижения размерности, визуализации как задача обучения без учителя.</p> <p><b>3.2. Метод центров тяжести (k-means)</b></p> <p>Использование метода центров тяжести для деления выборки на заданное количество кластеров <math>K</math> путем выбора их центров. Поиск центров кластеров через минимизацию суммарного расстояния от каждой точки до ближайшего центра с помощью метода локальной оптимизации. Ключевая роль начальной инициализации центров кластеров. Распространенным методом является случайный (с равной вероятностью) выбор <math>K</math> точек из имеющегося множества.</p> <p>Метод центров тяжести с евклидовой метрикой.</p> <p>Программная реализация.</p> <p>Разработка приложения для решения задач классификации.</p> <p>Функция вычисления ошибки классификации.</p> <p><b>3.3. Метод медиан</b></p> <p>Возможность сохранения и загрузки обученной модели в приложение для решения задач классификации.</p> <p>Реализация функции вычисления матрицы ошибок классификации.</p> <p>Реализация метода перекрестного контроля для подбора параметров алгоритмов обучения.</p>   |

#### 4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

| № п/п            | Темы практических занятий           | Формы практических занятий | Трудоемкость, (час) | Из них практической подготовки, (час) | № раздела дисциплины |
|------------------|-------------------------------------|----------------------------|---------------------|---------------------------------------|----------------------|
| <b>Семестр 6</b> |                                     |                            |                     |                                       |                      |
| 1                | Машинное обучение: предмет, задачи, | Решение ситуационных задач | 8                   | 8                                     | 1                    |



|       |                                       |                                   |    |   |   |
|-------|---------------------------------------|-----------------------------------|----|---|---|
|       | методы. История вопроса.              |                                   |    |   |   |
| 2     | Постановка задачи обучения с учителем | Решение ситуационных задач        | 4  | 4 | 2 |
| 3     | Метод к ближайших соседей             | Решение ситуационных задач        | 8  | 8 | 2 |
| 4     | Машина опорных векторов               | Решение ситуационных задач        | 8  | 8 | 2 |
| 5     | Деревья решений                       | Решение ситуационных задач        | 6  | 6 | 2 |
| 6     | Случайный лес                         | Решение ситуационных задач        | 8  | 8 | 2 |
| 7     | Градиентный бустинг деревьев решений  | Решение ситуационных задач        | 8  | 8 | 2 |
| 8     | Кластеризация                         | Решение ситуационных задач        | 8  | 8 | 3 |
| 9     | Метод центров тяжести                 | Решение ситуационных задач        | 6  | 6 | 3 |
| 10    | Метод медиан                          | Программная реализация приложения | 8  | 8 | 3 |
| Всего |                                       |                                   | 68 |   |   |

#### 4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

| № п/п                           | Наименование лабораторных работ | Трудоемкость, (час) | Из них практической подготовки, (час) | № раздела дисциплины |
|---------------------------------|---------------------------------|---------------------|---------------------------------------|----------------------|
| Учебным планом не предусмотрено |                                 |                     |                                       |                      |
|                                 |                                 |                     |                                       |                      |
| Всего                           |                                 |                     |                                       |                      |

#### 4.5. Курсовое проектирование/ выполнение курсовой работы

Цель курсовой работы:

Часов практической подготовки:

Примерные темы заданий на курсовую работу приведены в разделе 10 РПД.

#### 4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

| Вид самостоятельной работы                        | Всего, час | Семестр 6, час |
|---|------------|----------------|
| 1   | 2          | 3              |
| Изучение теоретического материала дисциплины (ТО) | 20         | 20             |
| Курсовое проектирование (КП, КР)                  |            |                |
| Расчетно-графические задания (РГЗ)                |            |                |
| Выполнение реферата (Р)                           |            |                |
| Подготовка к текущему контролю                    | 2          | 2              |

|  |    |    |
|--|----|----|
| успеваемости (ТКУ)                         |    |    |
| Домашнее задание (ДЗ)                      |    |    |
| Контрольные работы заочников (КРЗ)         |    |    |
| Подготовка к промежуточной аттестации (ПА) | 3  | 3  |
| Всего:                                     | 25 | 25 |

5. Перечень учебно-методического обеспечения  
для самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7–11.

6. Перечень печатных и электронных учебных изданий

Перечень печатных и электронных учебных изданий приведен в таблице 8.

Таблица 8– Перечень печатных и электронных учебных изданий

| Шифр/<br>URL адрес  | Библиографическая ссылка  | Количество экземпляров в библиотеке<br>(кроме электронных экземпляров) |
|---|---|--|
|   | Барсегян, А. А., Куприянов М. С., Степаненко В.В., Холод А.А., Методы и модели анализа данных: OLAP и Data Mining. БХВ-Петербург, 2004. - 336 с.  |  |
|   | Шалев-Шварц Ш. , Бен-ДавидШ. Идеи машинногообучения, ДМК Пресс, 2019<br><a href="https://e.lanbook.com/book/131686">https://e.lanbook.com/book/131686</a>   |  |
|   | <a href="#">Себастьян</a> Рашка, А. В. <a href="#">Логунов</a><br>Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения, ДМК Пресс, 2017.<br><a href="https://znanium.com/catalog/document?id=341047">https://znanium.com/catalog/document?id=341047</a> |  |
|   | <a href="#">Коэльо</a> Луис Педро, <a href="#">Вилли Ричарт</a> , <a href="#">Слинкин А. А.</a><br>Построение систем машинного обучения на языке Python, ДМК Пресс, 2016.<br><a href="https://znanium.com/catalog/document?id=341080">https://znanium.com/catalog/document?id=341080</a>  |  |
|   | В. В. Вьюгин Математические основы машинного обучения и прогнозирования, МЦНМО, 2014,<br><a href="https://znanium.com/catalog/document?id=328779">https://znanium.com/catalog/document?id=328779</a>  |  |
| ЭБС "ИНФРА-М"<br><a href="http://www.znanium.com">http://www.znanium.com</a><br><br>ЭБС «Лань»<br><a href="http://www.elan.com">http://www.elan.com</a> | 1. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufman, 2006. (Глава 1)<br><br>2. Tom Mitchell, Machine Learning, McGraw Hill, 1997. (Глава 1)<br><br>3. Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2   |  |

|  |  |  |
|--|--|--|
|  | <p>editions, 2005. (Главы 1, 2, 3)</p> <p>4. Биркгоф Г. Теория решеток. - М.: Наука, 1984.</p> <p>5. Szathmary, L. Symbolic Data Mining Methods with the Coron Platform. PhD thesis, University Henri Poincaré -- Nancy 1, France, Nov 2006. (глава The Coron User Tooolkit)</p> <p>6. Janez Demsar, Blaz Zupan, Gregor Leban, Tomaz Curk: Orange: From Experimental Machine Learning to Interactive Data Mining. J.-F. Boulicaut et al. (Eds.): PKDD 2004, LNAI 3202, pp. 537–539, Springer, 2004</p> <p>7. Николенко, С.И., Тулупьев, А.Л., Самообучающиеся системы. – М.: МЦНМО, 2009. – 288 с. (Глава 6)</p> <p>8. L. Zhukov, Spectral Clustering of Large Advertiser Datasets. Technical report, Overture R&amp;D, 2003.</p> <p>9. Автоматическое порождение гипотез в интеллектуальных системах/ Сост. Е.С. Панкратова, В.К. Финн; Под общ. ред. В.К.Финна. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 528 с. (Главы 1,2)</p> |  |
|--|--|--|

7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

| URL адрес   | Наименование                                     |
|---|--|
| <a href="http://www.machinelearning.ru/wiki/index.php?title=ML">http://www.machinelearning.ru/wiki/index.php?title=ML</a> | Машинное обучение (курс лекций, К.В.Воронцов)    |
| <a href="http://www.machinelearning.ru/wiki/index.php?title=29">http://www.machinelearning.ru/wiki/index.php?title=29</a> | Машинное обучение (семинары, ВМК МГУ)            |
| <a href="http://www.machinelearning.ru/wiki/index.php?title=29">http://www.machinelearning.ru/wiki/index.php?title=29</a> | Машинное обучение (курс лекций, Н.Ю.Золотых)     |
| <a href="http://www.machinelearning.ru/wiki/index.php?title=29">http://www.machinelearning.ru/wiki/index.php?title=29</a> | Машинное обучение (курс лекций, СГАУ, С.Лисицын) |

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

| № п/п | Наименование  |
|-------|---|
| 1     | <p>MS Office 2013:<br/>           Номер лицензии: 62042637<br/>           Родительская программа: 92010750ZZE1506<br/>           Договор на приобретение: 402-7</p> |

|   |   |
|---|---|
| 2 | Офис:<br>Microsoft® Office Professional Plus 2007 Russian Тип лицензии: Academic<br>Номер лицензии 44260430 |
|---|---|

8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

| № п/п | Наименование  |
|-------|---|
| 1.    | Электронно-библиотечная система Лань URL: <a href="https://e.lanbook.com/">https://e.lanbook.com/</a>   |
| 2.    | Электронно-библиотечная система Znanium URL: <a href="https://znanium.com/">https://znanium.com/</a>    |
| 3.    | Информационно-правовой портал ГАРАНТ.РУ URL: <a href="http://www.garant.ru/">http://www.garant.ru/</a>  |
| 4.    | Правовая поддержка КОНСУЛЬТАНТПЛЮС URL: <a href="http://www.consultant.ru">http://www.consultant.ru</a> |

## 9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

| № п/п | Наименование составной части материально-технической базы  | Номер аудитории (при необходимости)   |
|-------|--|---|
| 1     | Аудитория для практических занятий<br>Компьютерный класс   | Аудитория укомплектована специализированной мебелью, техническими средствами обучения, служащими для представления учебной информации большой аудитории<br>Ауд. 34-10   |
| 2     | Аудитории общего пользования (для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации) | Аудитория укомплектована специализированной мебелью, техническими средствами обучения, служащими для предоставления учебной информации большой аудитории  |
| 3     | Библиотека, Интернет-класс ГУАП (для самостоятельной работы)   | Помещения укомплектованы специализированной мебелью, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечено доступом в электронную информационно-образовательную среду ГУАП |

## 10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

| Вид промежуточной аттестации | Перечень оценочных средств   |
|------------------------------|--|
| Экзамен                      | Список вопросов к экзамену;<br>Экзаменационные билеты;<br>Расчетные задачи.        |
| Выполнение курсовой работы   | Экспертная оценка на основе требований к содержанию курсовой работы по дисциплине. |

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 –Критерии оценки уровня сформированности компетенций

| Оценка компетенции                    | Характеристика сформированных компетенций   |
|---------------------------------------|---|
| 5-балльная шкала                      |   |
| «отлично»<br>«зачтено»                | <ul style="list-style-type: none"> <li>– обучающийся глубоко и всесторонне усвоил программный материал;</li> <li>– уверенно, логично, последовательно и грамотно его излагает;</li> <li>– опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления;</li> <li>– умело обосновывает и аргументирует выдвигаемые им идеи;</li> <li>– делает выводы и обобщения;</li> <li>– свободно владеет системой специализированных понятий.</li> </ul> |
| «хорошо»<br>«зачтено»                 | <ul style="list-style-type: none"> <li>– обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы;</li> <li>– не допускает существенных неточностей;</li> <li>– увязывает усвоенные знания с практической деятельностью направления;</li> <li>– аргументирует научные положения;</li> <li>– делает выводы и обобщения;</li> <li>– владеет системой специализированных понятий.</li> </ul>  |
| «удовлетворительно»<br>«зачтено»      | <ul style="list-style-type: none"> <li>– обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы;</li> <li>– допускает несущественные ошибки и неточности;</li> <li>– испытывает затруднения в практическом применении знаний направления;</li> <li>– слабо аргументирует научные положения;</li> <li>– затрудняется в формулировании выводов и обобщений;</li> <li>– частично владеет системой специализированных понятий.</li> </ul>                 |
| «неудовлетворительно»<br>«не зачтено» | <ul style="list-style-type: none"> <li>– обучающийся не усвоил значительной части программного материала;</li> <li>– допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении;</li> <li>– испытывает трудности в практическом применении знаний;</li> <li>– не может аргументировать научные положения;</li> <li>– не формулирует выводов и обобщений.</li> </ul>   |

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

| № п/п | Перечень вопросов (задач) для экзамена   | Код индикатора |
|-------|--|----------------|
| 1     | Анализ формальных понятий в контексте разработки данных. Примеры практических задач. | ПК-1. 3.1      |
| 2     | Поиск частых (замкнутых) множеств признаков. Алгоритм Argiori.                       | ПК-1. В.1      |
| 3     | Поиск ассоциативных правил. Примеры практических                                     | ПК-5. 3.1      |

|   |  |           |
|---|--|-----------|
|   | задач.<br>Байесовский подход в задачах классификации.<br>Классификация текстов и фильтрация спама.   |           |
| 4 | Деревья решений. Алгоритмы ID3 и C4.5. Примеры практических задач.<br>Меры и методы оценки качества результатов в разработке данных и машинном обучении. | ПК-5. У.1 |
| 5 | Кластеризация. Метод К-средних. Агломеративная кластеризация. Примеры практических задач.  | ПК-6. 3.1 |
| 6 | Спектральная кластеризация. Примеры практических задач.<br>Неточные множества. Основные определения и прикладное значение.                               | ПК-7. У.1 |
| 7 | Анализ формальных понятий в контексте разработки данных. Примеры практических задач.<br>ДСМ-метод в терминах АФП. Прикладные задачи.                     | ПК-8. 3.1 |

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

| № п/п | Перечень вопросов (задач) для зачета / дифф. зачета | Код индикатора |
|-------|---|----------------|
|       | Учебным планом не предусмотрено                     |                |

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

| № п/п | Примерный перечень тем для курсового проектирования/выполнения курсовой работы  |
|-------|---|
| 1     | Морфологический анализ текстов при помощи машинного обучения (на материале одного или нескольких языков)  |
| 2     | Система разработки данных и машинного обучения с помощью программных средств <b>Анализа Формальных Понятий</b>  |
| 3     | Частичные порядки и решетки. Диаграммы порядка. Объектно-признаковые данные и формальные контексты. Операторы Галуа. Решетки формальных понятий. Признаковые импликации |
| 4     | Concept Explorer – программная система анализа данных на основе АФП.  |
| 5     | Согон – система поиска частых множеств признаков и ассоциативных правил.  |
| 6     | Информационная энтропия и прирост информации. Алгоритмы ID3 и C4.5.   |
| 7     | Критерии остановки и отсечения.   |
| 8     | Меры и методы оценки качества обучения (скользящий контроль).   |
| 9     | Системы машинного обучения WEKA и Orange и ее средства для работы с деревьями решений. Средства оценки качества классификации в этих системах.                          |
| 10    | Методы кластеризации в системах Weka и Orange.  |
| 11    | Спектральная кластеризация средствами пакета MatLab и библиотеки NumPy для языка Python.  |
| 12    | Rough Set Exploration System (RSES) – программная система для исследования объектно-признаковых данных на основе неточных   |

|    |   |
|----|---|
|    | множеств.   |
| 13 | QUДА – система разработки данных и машинного обучения. ДСМ-метод в системе QUДА                                   |
| 14 | Наивная байесовская классификация в системе Orange. Реализация наивной байесовской классификации на языке Python. |

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

| № п/п | Примерный перечень вопросов для тестов | Код индикатора |
|-------|--|----------------|
|       | Не предусмотрено                       |                |

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

| № п/п | Перечень контрольных работ |
|-------|----------------------------|
|       | Не предусмотрено           |

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

## 11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала.

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

### Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

- Постановка проблемы;
- Обзор предметной области;
- Обзор литературы;
- Разработка проекта решения;
- Реализация и анализ результатов;
- Дальнейшие направления исследований.

11.2. Методические указания для обучающихся по прохождению практических занятий.

Практическое занятие является одной из основных форм организации учебного процесса, заключающаяся в выполнении обучающимися под руководством преподавателя комплекса учебных заданий с целью усвоения научно-теоретических основ учебной дисциплины, приобретения умений и навыков, опыта творческой деятельности.

Целью практического занятия для обучающегося является привитие обучающимся умений и навыков практической деятельности по изучаемой дисциплине.

Планируемые результаты при освоении обучающимся практических занятий:

- закрепление, углубление, расширение и детализация знаний при решении конкретных задач;
- развитие познавательных способностей, самостоятельности мышления, творческой активности;
- овладение новыми методами и методиками изучения конкретной учебной дисциплины;
- выработка способности логического осмысления полученных знаний для выполнения заданий;
- обеспечение рационального сочетания коллективной и индивидуальной форм обучения.

Требования к проведению практических занятий

**Методические материалы преподавателю и студентам:**

Практическая часть курса организована в виде лабораторного практикума, состоящего из 7 лабораторных работ. Каждая лабораторная работа имеет типовую структуру.

Лабораторная работа №1

**Исследование объектно-признаковых данных с помощью программных средств  
Анализа Формальных Понятий**

Цель работы:

Необходимое программное обеспечение:

Теоретические сведения

Краткое описание ПО

Вопросы для допуска к работе

Задания для модельных расчетов

Задания для выполнения работы

Требования к отчету

Контрольные вопросы

Основная литература

Дополнительная литература

Успешное выполнение лабораторной работы предполагает написание студентом типового отчета. Отчет обязан содержать следующие элементы:



1. Титульный лист с указанием ФИО и номера группы, номера лабораторной работы.
  2. Цель работы.
  3. Ход работы с описанием выполнения лабораторной работы по заданиям. Необходимо включать таблицы результатов, промежуточные выводы и графики.
  4. Выводы. Как минимум работа должна содержать три нетривиальных вывода.
- Отчет о выполнении лабораторной работы проверяется преподавателем и предполагает его защиту в устной или письменной форме.

Необходимое для выполнения работ программное обеспечение, как правило, находится в свободном доступе и его можно загрузить в сети Интернет или скопировать с диска, прилагающегося к курсу. В лабораторных работах курса используются следующие открытые программные системы:

- Weka 3 – Data Mining Software in Java (разработана командой специалистов Университета Вайкато, Новая Зеландия); <http://www.cs.waikato.ac.nz/ml/weka/>
- Orange – Data Mining Fruitful & Fun (пакет создан лабораторией искусственного интеллекта Университета Любляны, Словения); <http://www.aillab.si/orange/>
- QuDA – Data Miner’s Discovery Environment (разработана в техническом Университете города Дармштадта, Германия); <http://sourceforge.net/projects/quda/>
- Coron System – платформа раскопок данных (разработана коллегами из группы Orpailleur в лаборатории LORIA Университета Нанси, Франция); <http://coron.loria.fr/>
- Concept Explorer – один из основных инструментов анализа формальных понятий (разработана в техническом Университете города Дармштадта, Германия); <http://conexp.sourceforge.net/>
- RSES2 – Rough Set Exploration System (разработана в институте математики Университета Варшавы, Польша). <http://logic.mimuw.edu.pl/~rses/>

### **Примеры вопросов и расчетных задач для допуска к сдаче экзамена.**

#### **Задачи**

1. Для заданного формального контекста:

|   | a | b | c | d |
|---|---|---|---|---|
| 1 | x | x |   | x |
| 2 |   |   | x |   |
| 3 |   | x |   | x |
| 4 | x |   |   | x |
| 5 | x | x | x |   |

- а) найдите множество формальных понятий;
- б) постройте диаграмму решётки формальных понятий;
- в) приведите примеры трех нетривиальных импликаций.

2. По заданной объектно-признаковой таблице с помощью алгоритма Apriori найдите все частые множества признаков при значении минимальной поддержки  $minsupp=0,5$ .

|   | a | b | c | d |
|---|---|---|---|---|
| 1 | x |   | x | x |
| 2 |   | x |   | x |
| 3 | x | x |   |   |
| 4 | x |   | x | x |
| 5 |   | x | x |   |

3. По объектно-признаковой таблице найдите множество ассоциативных правил при значении минимальной поддержки  $minsupp=0,5$  и минимальной достоверности  $minconf=0,6$ .

|   | a | b | c | d |
|---|---|---|---|---|
| 1 |   | x | x |   |
| 2 | x |   |   | x |
| 3 | x |   | x |   |
| 4 |   | x |   | x |
| 5 |   | x | x | x |

4. Для заданной объектно-признаковой таблицы о результатах футбольных матчей с целевым признаком «Победа» постройте дерево решений и классифицируйте объект матч, значение целевого признака для которого неизвестно.
- Для заданной объектно-признаковой таблицы с помощью алгоритма k-средних постройте 3 кластера, используя расстояние Евклида.
  - Для заданной таблицы решений найдите нижнее и верхнее приближение целевого класса «Играть?» со значениями «Да» и «Нет», т.е. для множеств  $X=\{1,2,6,8,14\}$  и  $Y=\{3,4,5,7,9,10,11,12,13\}$ , на следующих подмножествах признаков:

$$B1 = \{\text{«Температура», «Влажность»}\}$$

$$B2 = \{\text{«Облачность», «Ветер»}\}$$

|  | Облачность | Температура | Влажность | Ветер | Играть? |
|--|------------|-------------|-----------|-------|---------|
|  |            |             |           |       |         |

|    |          |          |            |         |     |
|----|----------|----------|------------|---------|-----|
| 1  | Солнечно | Жарко    | Высокая    | Слабый  | Нет |
| 2  | Солнечно | Жарко    | Высокая    | Сильный | Нет |
| 3  | Облачно  | Жарко    | Высокая    | Слабый  | Да  |
| 4  | Дождь    | Умеренно | Высокая    | Слабый  | Да  |
| 5  | Дождь    | Холодно  | Нормальная | Слабый  | Да  |
| 6  | Дождь    | Холодно  | Нормальная | Сильный | Нет |
| 7  | Облачно  | Холодно  | Нормальная | Сильный | Да  |
| 8  | Солнечно | Умеренно | Высокая    | Слабый  | Нет |
| 9  | Солнечно | Холодно  | Нормальная | Слабый  | Да  |
| 10 | Дождь    | Умеренно | Нормальная | Слабый  | Да  |
| 11 | Солнечно | Умеренно | Нормальная | Сильный | Да  |
| 12 | Облачно  | Умеренно | Высокая    | Сильный | Да  |
| 13 | Облачно  | Жарко    | Нормальная | Слабый  | Да  |
| 14 | Дождь    | Умеренно | Высокая    | Сильный | Нет |

5. Для таблицы решений  $(U, A, d)$  подмножество  $B \subseteq A$  называется решающим суперредуктом решения (decision super-reduct), если оно имеет нижнее и верхнее приближение целевого признака, как и само множество всех признаков  $A$ . Найдите все суперредукты решения.
1. Для заданного многозначного контекста найдите множество положительных и отрицательных гипотез. Предскажите значение целевого признака объектов ...
  2. Проведите классификацию объектов с неизвестным значением целевого признака, используя метод Naïve Bayes.

### Вопросы

1. Опишите и объясните способы вычисления расстояния между объектами в задачах кластеризации?
  1. Опишите и объясните способы измерения расстояний между кластерами.
  2. Объясните идею метода K-средних.
  3. Объясните работу алгоритма Apriori.
  4. Объясните принцип наивной классификации Байеса.
  5. Опишите меры оценки качества в машинном обучении и объясните работу метода скользящего контроля.

### Примеры заданий для контрольной работы.

1. Для многозначного контекста, заданного объектно-признаковой таблицей:
  - провести номинальное шкалирование признаков и найти формальные понятия для контекстов положительных и отрицательных примеров (целевой признак при шкалировании рассматривается только как пометка + и - примеров);
  - построить диаграммы решеток понятий положительного и отрицательного контекста;
  - найти максимально общие нефальсифицированные положительные и отрицательные гипотезы;
  - классифицировать недоопределенные примеры.

| Автоугон | Цвет    | Тип        | Производство | Повреждения | Угоняют? |
|----------|---------|------------|--------------|-------------|----------|
| 1        | Красный | Спортивный | США          | нет         | +        |
| 2        | Желтый  | Спортивный | Япония       | нет         | +        |
| 3        | Желтый  | Джип       | Япония       | нет         | +        |
| 4        | Красный | Спортивный | Япония       | есть        | +        |
| 5        | Желтый  | Спортивный | США          | есть        | -        |
| 6        | Желтый  | Джип       | США          | нет         | -        |
| 7        | Красный | Джип       | Япония       | есть        | -        |
| 8        | Зеленый | Спортивный | США          | нет         | τ        |
| 9        | Красный | Спортивный | Германия     | нет         | τ        |
| 10       | Черный  | Джип       | Япония       | нет         | τ        |

1. Для заданного контекста, найдите:
  - три нетривиальных импликации (импликация  $A \rightarrow B$  называется тривиальной, если  $B \subseteq A$ );
  - приведите примеры трех ассоциативных правил, не являющихся импликациями, укажите для них меры поддержки (support) и достоверности (confidence).
2. С помощью метода Naïve Bayes предскажите целевой класс для объектов 8–10 из 10 задачи.

11.3. Методические указания для обучающихся по прохождению курсового проектирования/выполнения курсовой работы.

Курсовой проект/ работа проводится с целью формирования у обучающихся опыта комплексного решения конкретных задач профессиональной деятельности.

Курсовой проект/ работа позволяет обучающемуся:

#### Структура пояснительной записки курсового проекта/ работы

Курсовая работа по дисциплине «Машинное обучение» предполагает следующую структуру: введение, две главы (в теоретической главе должен быть представлен реферативный обзор научной литературы по теме курсовой работы, в практической части работы должны быть представлены результаты собственного перевода), заключения. В пояснительной записке к курсовой работе должна быть указана дисциплина, по которой защищается курсовая работа, тема работы, ФИО руководителя, ФИО студента, номер группы.

#### Требования к оформлению пояснительной записки курсового проекта/ работы

Пояснительная записка должна быть оформлена в соответствии с требованиями СТО ГУАП.СМК 3.170

#### Методические рекомендации по программной реализации приложений

### **1. Разработка приложения для решения задач классификации**

#### **1.1. Требования к приложению**

В рамках данной лабораторной работы предлагается разработать приложение для решения задач классификации, предусматривающее возможность применения рассмотренных выше алгоритмов машинного обучения. К приложению предъявляются следующие требования:

1. Организация диалога с пользователем для загрузки набора данных из файла с последующим выбором алгоритма обучения с учителем и его параметров.
2. Обучение модели на загруженных данных.
3. Вычисление ошибки классификации на обучающей и тестовой выборках.
4. Визуализация результата работы алгоритмов в случае двумерного пространства признаков.

#### **1.2. Структура приложения**

Приложение будет состоять из набора модулей (cvsvm.cpp/h, cvdtree.cpp/h, cvrtrees.cpp/h, cvgbtrees.cpp/h), каждый из которых предназначен для работы (запрос у пользователя параметров, обучение и предсказание) рассматриваемых алгоритмов обучения, модуля вычисления ошибки классификации (errorMetrics.cpp/h), модуля визуализации данных (drawingFunctions.cpp/h) и основного модуля (main.cpp), содержащего общую логику работы программы. Основной модуль и модуль визуализации предоставляются в реализованном виде, следовательно, необходимо реализовать непосредственно работу с различными алгоритмами обучения с учителем. Далее рассматриваются интерфейсы готовых функций, а также функционал, который предлагается реализовать самостоятельно.

Основная логика программы сосредоточена в модуле main.cpp и выглядит следующим образом:

- В цикле пользователю предлагается загрузить данные из файла, либо применить к уже загруженным данным алгоритм обучения с учителем из списка: машина опорных векторов, дерево решений, случайный лес, градиентный бустинг деревьев решений.
- Если выбран алгоритм решения задачи классификации, у пользователя запрашиваются параметры алгоритма обучения, производится обучение и вычисление ошибок на обучающей и тестовой выборках. Если размерность пространства признаков равна двум, то данные и разбиение пространства признаков полученной моделью отображаются в новом графическом окне.

- Графические окна не обязательно закрывать перед применением другого алгоритма к тем же данным, что позволяет сравнить результаты работы различных методов. Однако при загрузке новых данных все графические окна будут закрыты автоматически.

## **2. Разработка приложения для решения задач кластеризации**

### **2.1. Требования к приложению**

Также в рамках данной лабораторной работы предлагается разработать приложение для кластеризации точек методом центров тяжести. К приложению предъявляются следующие требования:

1. Загрузка данных из файла, имя которого указывается в качестве параметра командной строки.
2. Выполнение кластеризации на заданное (в виде аргумента командной строки) число кластеров.
3. Визуализация кластеризации в двумерном пространстве.

### **2.2. Структура приложения**

Приложение будет состоять из двух модулей: основной (`main.cpp`) и модуль визуализации (`drawingFunctions.cpp/h`). Функции визуализации предоставляются в готовом виде и аналогичны описанным в разделе 1.2.

В основном модуле должна выполняться следующая последовательность действий:

1. Загрузка данных для кластеризации из файла.
2. Выполнение кластеризации методом центров тяжести на указанное количество кластеров.
3. Визуализация результата.

Код, необходимый для загрузки и визуализации предоставляется в готовом виде, следовательно, требуется лишь написать вызов функции кластеризации `kmeans`.

После того, как код основного модуля будет дописан, предлагается запустить программу на предоставленных наборах данных (`dataset1.yml`, `dataset2.yml`, `dataset3.yml`, `dataset4.yml`) и проанализировать полученные результаты.

### **3. Дополнительные задания**

1. Реализуйте возможность сохранения и загрузки обученной модели в приложении для решения задач классификации.
2. Реализуйте функцию вычисления матрицы ошибок классификации  $\mathcal{E}$ , где элемент  $\mathcal{E}_{i,j}$  равен количеству прецедентов выборки принадлежащих к классу  $j$  и отнесенных алгоритмом классификации к классу  $i$ .
3. Реализуйте метод перекрестного контроля для подбора параметров алгоритмов обучения

## **11.4. Методические указания для обучающихся по прохождению самостоятельной работы**

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся, являются:

- учебно-методический материал по дисциплине.

Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику и тем самым проникнуть в творческую лабораторию автора.

Ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд подсобных материалов для быстрого повторения прочитанного, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

11.5. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Главное условие успешности в освоении учебной дисциплины - систематические занятия. Для полного понимания изучаемого материала следует задавать вопросы непосредственно на практических занятиях, чтобы не оставлять пробелов в изучении. За дополнительными разъяснениями и рекомендациями студент может обращаться к преподавателю во время консультаций. Систематическая работа в семестре, активное участие в семинарах (доклады, сообщения, самостоятельно подготовленные презентации по интересующим темам по теории и практике перевода) и выполнение практических заданий обеспечит высокую оценку при прохождении промежуточной аттестации.

11.6. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя:

– экзамен – форма оценки знаний, полученных обучающимся в процессе изучения всей дисциплины или ее части, навыков самостоятельной работы, способности применять их для решения практических задач. Экзамен, как правило, проводится в период экзаменационной сессии и завершается аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

При подготовке к аттестации следует не только разобраться в материале, но попробовать, не подглядывая в конспекты или учебники, изложить письменно наиболее существенные понятия, утверждения, точки зрения по каждому разделу программы, составить план-конспекты ответов на вопросы.

На экзамене и дифференцированном зачете в процессе подготовки к ответу прежде, чем приступить к подробному изложению ответа на вопрос, следует составить (письменно или устно) план предстоящего ответа, обязательно привести примеры, указать авторов, на которых ссылаются при ответе.

Промежуточная аттестация проводится не только в традиционном формате «вопрос-ответ», но и в форме дискуссии, в процессе которой определяется умение студента быстро мыслить, формулировать свой ответ при линейном развитии речи, владение устной и письменной версией официально-деловой нормы современного русского языка.

Лист внесения изменений в рабочую программу дисциплины

| Дата внесения изменений и дополнений.<br>Подпись внесшего изменения | Содержание изменений и дополнений | Дата и № протокола заседания кафедры | Подпись зав. кафедрой |
|---|-----------------------------------|--------------------------------------|-----------------------|
|   |                                   |                                      |                       |
|   |                                   |                                      |                       |
|   |                                   |                                      |                       |
|   |                                   |                                      |                       |
|   |                                   |                                      |                       |