МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «РГГУ»)

Институт лингвистики

УНЦ компьютерной лингвистики

Рабочая программа дисциплины

«Методы классификации и машинное обучение»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Магистерская программа: Фундаментальная и компьютерная лингвистика

Квалификация выпускника: магистр

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

Методы классификации и машинное обучение

Рабочая программа дисциплины

Составитель:

к.ф-м.н., доцент А.А.Сорокин

Ответственный редактор:

д. филол. н., профессор В.И.Подлесская

УТВЕРЖДЕНО
Протокол заседания УНЦ компьютерной лингвистики

№ 1 от «28» августа 2019г.

Оглавление

1. Пояснительная записка

1.1 Предмет

Предметом дисциплины (модуля) является изучение основных методам машинного обучения и текстовой классификации. Курс проходит параллельно с курсом «Основные алгоритмы лингвистического анализа. В курсе рассматриваются как математические основы методов машинного обучения и статистического анализа данных, так и детали их практического применения, в частности, подробно изучается библиотека scikit-learn, содержащая реализацию основных алгоритмов машинного обучения на языке Python. Особое внимание уделяется использованию методов машинного обучения при классификации текстов, а также в других задачах лингвистического анализа.

1.2 Цель и задачи курса

Курс направлен на решение следующих задач:

- познакомить обучающихся с основными алгоритмами машинного обучения, применяемыми для решения лингвистических задач, а также с программными продуктами, реализующими данные методы;
- познакомить магистрантов с основными задачами текстовой классификации (жанровая, тематическая, анализ тональности и т. д.) и кластеризации, а также с используемыми в них алгоритмами машинного обучения;
- познакомить магистрантов с математическими методами, лежащими в основе алгоритмов машинного обучения;
- научить магистрантов как предварительно выбирать алгоритм машинного обучения для решения для прикладных лингвистических задач, так и дорабатывать выбранный алгоритм в зависимости от специфики задачи и исходных данных;
- научит магистрантов квалифицированно подбирать признаковое представление данных для алгоритмов машинного обучения, отражающее лингвистическую специфику задачи.
- научить магистрантов анализировать результаты применения статистических алгоритмов к лингвистическим данным;
- у магистрантов знания, позволяющие им квалифицированно читать литературу по специальности, включающую в себя как учебные материалы и научные статьи, так и более специализированные технические материалы, например, программную документацию.

1.3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способностью адаптироваться к новым теориям и результатам мировой науки и расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук (ОПК-6);

способностью выбирать оптимальные теоретические подходы и методы решения

конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);

способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);

способностью разрабатывать и совершенствовать системы автоматизации и информационной поддержки лингвистических исследований (ПК-10);

способностью производить систематизацию произвольной предметной области, разрабатывать для нее классификаторы, рубрикаторы, онтологии, проводить типологизацию данных и моделирование предметной области, разрабатывать универсальные онтологии (ПК-14)

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать:

- структуру научно-практической области исследований «машинное обучение» и ее место в контексте смежных наук, в частности, математики;
- основные задачи машинного обучения (обучение с учителем, без учителя, частичное обучение с учителем), а также алгоритмы, применяемые для решения данных задач;
- основные задачи текстовой классификации, а также их формальное описание в терминах машинного обучения;
- математические основы базовых алгоритмов машинного обучения;
- основные типы лингвистических ресурсов, используемых для получения исходных данных, которые впоследствии применяются в алгоритмах машинного обучения;

Уметь:

- локализовать практическую задачу в контексте организации научно-практической области исследований «машинное обучение» и находить средства для ее решения;
- самостоятельно подбирать базовый алгоритм машинного обучения для решения той или иной задачи прикладной лингвистики, а также обосновывать его выбор;
- представлять лингвистические данные в виде, который может быть подан на вход алгоритмам машинного обучения, а также обосновывать выбор того или иного представления;
- анализировать результаты работы алгоритмов машинного обучения и подбирать оптимальные параметры алгоритма как результат этой оценки;
- модифицировать выбранный алгоритм в зависимости от результатов его работы
- реализовывать выбранный алгоритм машинного обучения на одном из высокоуровневых языков программирования или пользоваться имеющимися реализациями, при необходимости внося изменения;
- пользоваться библиотекой scikit-learn для алгоритмов машинного обучения на языке Python;

Владеть:

- основными методами обработки статистического анализа данных.

1.4 Место дисциплины в структуре образовательной программы

Дисциплина (модуль) «Методы классификации и машинное обучение» является

дисциплиной по выбору вариативной части цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 2 курса (3 семестр). Дисциплина (модуль) реализуется кафедрой компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение *домашних заданий; лабораторные работы; защита исследовательского проекта*; промежуточная аттестация в форме: *зачет.*

Общая трудоемкость освоения дисциплины (модуля) составляет *2* зачетные единицы, *72* часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – *20 часов*; самостоятельная работа студента – *52 часа*.

2. Структура дисциплины

№ п/п	**Раздел Дисциплины**	**Семестр**	**Неделя семестра**	**Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)**				**Формы текущего контроля успеваемости *(по неделям семестра)* Форма промежуточной аттестации *(по семестрам)***
				лекции	семинары	самостоятельная работа		
1.	Введение. Основные задачи машинного обучения, их применение в компьютерной лингвистике.	1	1			10		
2.	Задачи классификации в обработке естественного языка и компьютерной лингвистике. Признаковое описание объектов, типы признаков	1	2		2	10		ДЗ1. Подбор признакового описания для выбранной задачи компьютерной лингвистики.
3.	Векторная модель классификации, матрица объекты-признаки, линейные классификаторы.	1	3		2	8		
4.	Библиотека scikit-learn для машинного обучения. Основные	1	4		2	4		ДЗ2. Знакомство с библиотекой scikit-learn.

	объекты и типы данных.							
5.	Наивный байесовский классификатор, вероятностная модель. Его недостатки и достоинства..	1	5		2	4	4	ДЗ3. Тестирование наивного байесовского классификатора на модельной задаче.
6.	Линейные классификаторы в библиотеке scikit-learn, основные алгоритмы.	3	6			6		ДЗ4. Исследовательский проект (часть 1): автоматическое определение языка коротких текстов.
7.	Извлечение признаков из текста, меры качества признаков. Оценка качества классификации.	1	7		4	6		ДЗ5. Исследовательский проект (часть 2): автоматическое определение языка коротких текстов.
8.	Автоматическая классификация текстов, модель мешка слов. Стандартные приёмы в задачах текстовой классификации	1	8		4	4		ДЗ6. Классификация текстов по тематическим категориям.
9.		1	9				2	Проверка и обсуждение исследовательского проекта и ДЗ6.
10.	Обучение без учителя, кластеризация, стандартные алгоритмы: иерархическая кластеризация и метод средних.	1	10			6		ДЗ7. Кластеризация «вручную» набора данных различными методами.
11.	Дистрибутивная семантика, семантические вектора, их применение в различных задачах.		11		4	4		ДЗ8. Применение семантических векторов для автоматической классификации.
12.	Дистрибутивная семантика на основе нейронных сетей. Обзор её применений.	1	12		2	4		
13.	Обзор применений машинного обучения в различных задачах компьютерной лингвистики (морфология, синтаксис и т.д.)	1	13			2		ДЗ9. Применение машинного обучения в автоматическом морфологическом анализе.
14.	Зачет	1	14			10		Контрольные вопросы
	Итого:				20	52		

3. Содержание дисциплины

1. Основные задачи машинного обучения, их применение в компьютерной лингвистике.

Общая постановка задачи машинного обучения, обучение с учителем и без учителя. Примеры задач машинного обучения. Основные задачи компьютерной лингвистики. Их интерпретация как задач машинного обучения.

2. Задачи классификации в обработке естественного языка и компьютерной лингвистике. Признаковое описание объектов, типы признаков

Общая постановка задачи классификации. Типы признаков: двоичные, категориальные, вещественные, порядковые. Двоичная и многоклассовая классификация. Задачи компьютерной лингвистики как задачи классификации. Типы признаков в различных задачах, их извлечение из текста.

3. Векторная модель классификации, матрица объекты-признаки, линейные классификаторы.

Векторное представление текста в задачах компьютерной лингвистики, модель мешка слов. Матрица объекты-признаки. Сведение задачи классификации к подбору оптимальной разделяющей плоскости. Свойства задач текстовой классификации (большое количество признаков, разреженность).

4. Библиотека scikit-learn для машинного обучения. Основные объекты и типы данных.

Решение задач машинного обучение на языке Python. Представление данных, модуль Numpy, массивы и матрицы. Операции с матрицами и векторами в модуле Numpy. Библиотека scikit-learn, основные классы и методы. Разреженные матрицы и модуль scipy.

5. Наивный байесовский классификатор, вероятностная модель.

Вероятностная модель классификации, наивный байесовский классификатор. Его связь с моделью мешка слов. Достоинства и недостатки наивного байесовского классификатора. Проблема нулевых вероятностей, сглаживание признаков. Реализация наивного байесовского классификатора в библиотеке scikit-learn.

6. Линейные классификаторы в библиотеке scikit-learn, основные алгоритмы.

Понятие линейного классификатора, случай двух и более классов. Решающее правило классификатора, вектор весов, разделяющая гиперплоскость. Основные алгоритмы нахождения вектора весов: метод опорных векторов, логистическая регрессия.

7. Извлечение признаков из текста, меры качества признаков. Оценка качества классификации.
Стандартные признаки в задачах классификации: подсчёт количество слов, символьные энграммы. Зависимость качества классификации от используемых признаков. Меры оценки качества (точность, полнота, F-мера). Обучающая и контрольная выборка, скользящий контроль, переобучение. Отбор признаков, основные методы оценки качества признака (вес признака, вероятность класса).

8. Обучение без учителя, кластеризация, стандартные алгоритмы: иерархическая кластеризация и метод средних.
Понятие обучения без учителя, постановка задачи кластеризации. Методы измерения расстояния между объектами, их зависимость от задачи. Иерархическая кластеризация, вычисление межкластерного расстояния. Метод k-средних, его сравнение с иерархической кластеризацией. Применения кластеризации.

9. Дистрибутивная семантика, семантические вектора, их применение в различных задачах.

Дистрибутивная семантика, основные подходы. Матрица совместных вхождений. Понижение размерности семантических векторов

10. Дистрибутивная семантика на основе нейронных сетей. Обзор её применений

Общее знакомство с нейронными сетями. Методы skipgram и CBOW для получения дистрибутивных векторов, модель Т. Миколова. Свойства дистрибутивных векторов. Применение дистрибутивных векторов в задачах компьютерной лингвистики.

11. Обзор применений машинного обучения в различных задачах компьютерной лингвистики (морфология, синтаксис и т.д.).

Применение методов машинного обучения в различных задачах компьютерной лингвистики. Автоматический морфологический анализ, определение грамматической категории слова на основе признаков. Автоматический синтаксический анализ, использование машинного обучения для снятия неоднозначности. Переранжирование гипотез.

4. Образовательные технологии

В соответствии с требованиями ФГОС по направлению 45.04.03 «Фундаментальная и прикладная лингвистика» и с учетом специфики магистерской программы «Фундаментальная и компьютерная лингвистика» занятия лекционного типа составляют не более 20% аудиторных занятий, а удельный вес занятий, проводимых в интерактивных формах, составляют не менее 40% аудиторных занятий. Интерактивные формы обучения в данном курсе предполагают:

1. систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с отчетом по результатам исследования);
2. он-лайн демонстрации работы с лингвистическими интернет-источниками (и др.);
3. он-лайн использование лингвистических ресурсов (Национальный корпус русского языка, Лексико-семантические базы и др.);
4. обсуждения курсовых исследовательских проектов;
5. работа в группах по выполнению домашних практических заданий.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

<table>
<tr><th>100-балльная шкала</th><th colspan="2">Традиционная шкала</th><th>Шкала ECTS</th></tr>
<tr><td>95 – 100</td><td rowspan="2">отлично</td><td rowspan="5">зачтено</td><td>A</td></tr>
<tr><td>83 – 94</td><td>B</td></tr>
<tr><td>68 – 82</td><td>хорошо</td><td>C</td></tr>
<tr><td>56 – 67</td><td rowspan="2">удовлетворительно</td><td>D</td></tr>
<tr><td>50 – 55</td><td>E</td></tr>
<tr><td>20 – 49</td><td rowspan="2">неудовлетворительно</td><td rowspan="2">не зачтено</td><td>FX</td></tr>
<tr><td>0 – 19</td><td>F</td></tr>
</table>

Распределение баллов по видам учебной деятельности таково:

- посещение семинарских занятий – до 8 баллов,
- уровень активности студента при подготовке к занятиям (конспектирование специальной литературы, готовность отвечать на вопросы по анализу кейсов, активное участие в дискуссиях, коллоквиумах и мозговом штурме и проч.) и во время проведения занятий (участие в обсуждениях и выполнении коллективных заданий) – всего до 32 баллов,
- качество выполнения контрольной работы (текущая аттестация) – до 20 баллов,
- успешность выполнения итогового творческого задания – до 40 баллов.

Оценка «зачтено» выставляется, если студент набрал в сумме не менее 50 баллов. Магистрант, не набравший в сумме 50 баллов, сдаёт зачёт по всему курсу и предъявляет преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

5.2. *Критерии выставления оценок*

При выставлении оценки преподаватель ориентируется на следующие содержательные критерии.

Количество баллов	Критерии оценки
95−100 (A)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Свободно ориентируется в учебной и профессиональной литературе. Компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».
83−94 (B)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, почти все задания, предусмотренные рабочей программой дисциплины, выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному. Обучающийся адекватно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Достаточно свободно ориентируется в учебной и профессиональной литературе. Почти все компетенции, закреплённые за дисциплиной,

Количество баллов	Критерии оценки
	сформированы на уровне «высокий».
68–82 (C)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе. Компетенции, закреплённые за дисциплиной, сформированы на уровне «хороший».
56–67 (D)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
50–55 (E)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные рабочей программой дисциплины учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».

Количество баллов	Критерии оценки
21–49 (FX)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.
0–20 (F)	Оценка выставляется с учетом текущей и промежуточной аттестации. Теоретическое содержание дисциплины не освоено. Необходимые практические навыки работы не сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены с грубыми ошибками. Дополнительная самостоятельная работа над материалом дисциплины не приведет к какому-либо значимому повышению качества выполнения учебных заданий. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Текущий контроль успеваемости студентов проводится в следующих формах: выполнение домашних заданий (10 заданий – 50 баллов максимум); тестовое задание (максимально 10 баллов); защита исследовательского проекта (максимально - 40 баллов). Для получения удовлетворительной оценки необходимо набрать минимум 60 баллов.

В качестве домашних заданий предлагаются задания следующих типов

ДЗ1. Подбор призна-кового описания для выбранной задачи компьютерной лингвистики.
ДЗ2. Знакомство с библиотекой scikit-learn.
ДЗ3. Тестирование наивного байесовского классификатора на модельной задаче.
ДЗ4. Исследовательский проект (часть 1): автоматическое определение языка коротких текстов.
ДЗ5. Исследовательский проект (часть 2): автоматическое определение языка коротких текстов.
ДЗ6. Классификация текстов по тематическим категориям.
Проверка и обсуждение исследовательского проекта и ДЗ6.
ДЗ7. Кластеризация «вручную» набора данных различными методами.
ДЗ8. Применение семантических векторов для автоматической классификации.
ДЗ9. Применение машинного обучения в автоматическом морфологическом анализе.

Зачет ориентирован на следующие контрольные вопросы

Основные типы задач машинного обучения.
Интерпретация задач компьютерной лингвистики в терминах машинного обучения.
Признаковое описание, типы признаков, их зависимость от задачи.
Наивный байесовский классификатор.
Линейные алгоритмы классификации.
Оценка качества классификации, основные меры.
Стандартное признаковое описание в задачах текстовой классификации.
Постановка задачи кластеризации, основные алгоритмы.
Дистрибутивная семантика, применение нейронных сетей.
Машинное обучение в задачах компьютерной морфологии и синтаксиса.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список литературы

Основная литература

1. Бахвалов Ю.Н., Малыгин Л.Л., Черкас П.С. Метод машинного обучения на основе алгоритма многомерной интерполяции и аппроксимации случайных функций. // Вестник Череповецкого государственного университета, 2012. – 4 стр.
2. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. – О
3. Reilly Media, 2017 – 340 с.
4. Sebastani F. machine Learning in Automated Text Categorization. 2001 – 77 p.
5. Scikit-learn, библиотека алгоритмов машинного обучения. http://scikit-learn.org/stable/
6. Ресурс по распознаванию данных и машинному обучению. http://machinelearning.ru

Рекомендованная литература

1. К. В. Воронцов. Лекции по машинному обучению. http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_(%D0%BA%D1%83%D1%80%D1%81_%D

0%BB%D0%B5%D0%BA%D1%86%D0%B8%D0%B9%2C_%D0%9A.%D0%92.%D0%92%D0%BE%D1%80%D0%BE%D0%BD%D1%86%D0%BE%D0%B2)
2. Прикладная и компьютерная лингвистика, Под. ред. А. В. Митрениной. М., УРСС, 2016.
3. Manning C. D. et al. Foundations of statistical natural language processing. – Cambridge : MIT press, 1999
4. Bishop C. M. Pattern recognition //Machine Learning. – 2006

Программное обеспечение и Интернет-ресурсы

Учебная дисциплина должна быть обеспечена учебно-методической документацией и материалами. Обязательная литература должна быть представлена в библиотеке ВУЗа, сети Интернет или локальной сети ВУЗа (факультета). Для обучающихся должен быть обеспечен онлайн доступ к Интернет источникам и системам. В частности, в процессе обучения используются следующие Интернет-ресурсы:

Интернет-источники
Прикладные ресурсы:

1. Scikit-learn, библиотека алгоритмов машинного обучения. http://scikit-learn.org/stable/
2. Ресурс по распознаванию данных и машинному обучению. http://machinelearning.ru

7. Материально-техническое обеспечение дисциплины

Занятия по курсу «Методы классификации и машинное обучение» можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимо также наличие доски, чтобы преподаватель мог разбирать примеры по ходу объяснения и записывать задания.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

При необходимости рабочая программа дисциплины может быть адаптирована для обеспечения образовательного процесса лицам с ограниченными возможностями здоровья, в том числе для дистанционного обучения. Для этого от студента требуется представить заключение психолого-медико-педагогической комиссии (ПМПК) и личное заявление (заявление законного представителя).

В заключении ПМПК должно быть прописано:

- рекомендуемая учебная нагрузка на обучающегося (количество дней в неделю, часов в день);
- оборудование технических условий (при необходимости);
- сопровождение и (или) присутствие родителей (законных представителей) во время учебного процесса (при необходимости);
- организация психолого-педагогического сопровождение обучающегося с указанием специалистов и допустимой нагрузки (количества часов в неделю).

Для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся, при необходимости могут быть созданы фонды оценочных средств, адаптированные для лиц с ограниченными возможностями здоровья и позволяющие оценить достижение ими запланированных в основной образовательной программе результатов обучения и уровень сформированности всех компетенций, заявленных в образовательной программе.

Форма проведения текущей и итоговой аттестации для лиц с ограниченными возможностями здоровья устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно (на бумаге, на компьютере), в форме тестирования и т.п.). При необходимости студенту предоставляется дополнительное время для подготовки ответа на зачете или экзамене.

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

- письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;

- обеспечивается индивидуальное равномерное освещение не менее 300 люкс;

- для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;

- письменные задания оформляются увеличенным шрифтом;

- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих:

- лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;

- письменные задания выполняются на компьютере в письменной форме;

- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

- письменные задания выполняются на компьютере со специализированным программным обеспечением;

- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:

- в печатной форме увеличенным шрифтом;
- в форме электронного документа;
- в форме аудиофайла.

• для глухих и слабослышащих:

- в печатной форме;
- в форме электронного документа.

• для обучающихся с нарушениями опорно-двигательного аппарата:

- в печатной форме;
- в форме электронного документа;
- в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

• для слепых и слабовидящих:

- устройством для сканирования и чтения с камерой SARA CE;
- дисплеем Брайля PAC Mate 20;
- принтером Брайля EmBraille ViewPlus;

• для глухих и слабослышащих:

- автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
- акустический усилитель и колонки;

• для обучающихся с нарушениями опорно-двигательного аппарата:

- передвижными, регулируемыми эргономическими партами СИ-1;
- компьютерной техникой со специальным программным обеспечением.

9. Приложения

Приложение 1. Аннотация дисциплины

Предметом дисциплины (модуля) является изучение основных методам машинного обучения и текстовой классификации. Курс проходит параллельно с курсом «Основные алгоритмы лингвистического анализа. В курсе рассматриваются как математические основы методов машинного обучения и статистического анализа данных, так и детали их практического применения, в частности, подробно изучается библиотека scikit-learn, содержащая реализацию основных алгоритмов машинного обучения на языке Python. Особое внимание уделяется использованию методов машинного обучения при классификации текстов, а также в других задачах лингвистического анализа.

Курс направлен на решение следующих задач:

- познакомить обучающихся с основными алгоритмами машинного обучения, применяемыми для решения лингвистических задач, а также с программными продуктами, реализующими данные методы;
- познакомить магистрантов с основными задачами текстовой классификации (жанровая, тематическая, анализ тональности и т. д.) и кластеризации, а также с используемыми в них алгоритмами машинного обучения;
- познакомить магистрантов с математическими методами, лежащими в основе алгоритмов машинного обучения;
- научить магистрантов как предварительно выбирать алгоритм машинного обучения для решения для прикладных лингвистических задач, так и дорабатывать выбранный алгоритм в зависимости от специфики задачи и исходных данных;
- научит магистрантов квалифицированно подбирать признаковое представление данных для алгоритмов машинного обучения, отражающее лингвистическую специфику задачи.
- научить магистрантов анализировать результаты применения статистических алгоритмов к лингвистическим данным;
- у магистрантов знания, позволяющие им квалифицированно читать литературу по специальности, включающую в себя как учебные материалы и научные статьи, так и более специализированные технические материалы, например, программную документацию.

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способностью адаптироваться к новым теориям и результатам мировой науки и расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук (ОПК-6);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);

способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);

способностью разрабатывать и совершенствовать системы автоматизации и информационной поддержки лингвистических исследований (ПК-10);

способностью производить систематизацию произвольной предметной области,

разрабатывать для нее классификаторы, рубрикаторы, онтологии, проводить типологизацию данных и моделирование предметной области, разрабатывать универсальные онтологии (ПК-14)

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать:

- структуру научно-практической области исследований «машинное обучение» и ее место в контексте смежных наук, в частности, математики;
- основные задачи машинного обучения (обучение с учителем, без учителя, частичное обучение с учителем), а также алгоритмы, применяемые для решения данных задач;
- основные задачи текстовой классификации, а также их формальное описание в терминах машинного обучения;
- математические основы базовых алгоритмов машинного обучения;
- основные типы лингвистических ресурсов, используемых для получения исходных данных, которые впоследствии применяются в алгоритмах машинного обучения;

Уметь:

- локализовать практическую задачу в контексте организации научно-практической области исследований «машинное обучение» и находить средства для ее решения;
- самостоятельно подбирать базовый алгоритм машинного обучения для решения той или иной задачи прикладной лингвистики, а также обосновывать его выбор;
- представлять лингвистические данные в виде, который может быть подан на вход алгоритмам машинного обучения, а также обосновывать выбор того или иного представления;
- анализировать результаты работы алгоритмов машинного обучения и подбирать оптимальные параметры алгоритма как результат этой оценки;
- модифицировать выбранный алгоритм в зависимости от результатов его работы
- реализовывать выбранный алгоритм машинного обучения на одном из высокоуровневых языков программирования или пользоваться имеющимися реализациями, при необходимости внося изменения;
- пользоваться библиотекой scikit-learn для алгоритмов машинного обучения на языке Python;

Владеть:

- основными методами обработки статистического анализа данных.

Дисциплина (модуль) «Методы классификации и машинное обучение» является дисциплиной по выбору вариативной части цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 2 курса (3 семестр). Дисциплина (модуль) реализуется кафедрой компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение *домашних заданий; лабораторные работы; защита исследовательского проекта*; промежуточная аттестация в форме: *зачет.*

Общая трудоемкость освоения дисциплины (модуля) составляет *2* зачетные единицы, *72* часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – *20 часов*; самостоятельная работа студента – *52 часа*.

Приложение 2. Лист изменений

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение к листу изменений №1	*31.08.2020г*	**1**

Приложение к листу изменений №1

1. Структура дисциплины (к п. 2 РПД на 2020)

Общая трудоёмкость дисциплины составляет 2 з.е., 76 ч., в том числе контактная работа обучающихся с преподавателем 20 ч., самостоятельная работа обучающихся 56 ч.

2. Образовательные технологии (к п.4 на 2020 г.)

В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

– видео-лекции**;**
– онлайн-лекции в режиме реального времени;
– электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
– системы для электронного тестирования;
– консультации с использованием телекоммуникационных средств.

3. Перечень БД и ИСС (к п. 6 на 2020 г.)

№п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru

4. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п	Наименование ПО	Производитель	Способ распространения (*лицензионное или свободно распространяемое*)

1	Microsoft Share Point 2010	Microsoft	лицензионное
2	Windows 10 Pro	Microsoft	лицензионное
3	Kaspersky Endpoint Security	Kaspersky	лицензионное
4	Microsoft Office 2016	Microsoft	лицензионное
5	Zoom	Zoom	лицензионное