

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет романо-германской филологии

УТВЕРЖДАЮ:
И.о. проректора по учебной
работе, качеству образования
– первый проректор

Хасуров Т.А.
подпись
« 11 » 2025 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.О.15 Машинное обучение в лингвистике

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) Прикладные языковые технологии в цифровых коммуникациях и публичном управлении

Форма обучения очная

Квалификация магистр

Краснодар 2025

Рабочая программа дисциплины Б1.О.15 МАШИННОЕ ОБУЧЕНИЕ В ЛИНГВИСТИКЕ составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Программу составила:
Рябченко Н. А., канд. полит. н., доцент



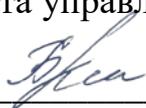
Рабочая учебная программа утверждена на заседании кафедры социальной работы, психологии и педагогики высшего образования (разработчика)
Протокол № 11 от «18» марта 2025 г.

Заведующий кафедрой СР и ППВО (разработчика)  Л.М. Чепелева

Утверждена на заседании учебно-методической комиссии факультета управления и психологии
протокол № 8 «20» марта 2025 г.

Председатель УМК факультета управления и психологии

Белокопытова К.М.
фамилия, инициалы


подпись

Рецензенты:

Савченко А.П., канд. физ.-мат. наук, доцент кафедры общего, стратегического, информационного менеджмента и бизнес-процессов, КубГУ

Воробьев Г.А., профессор, зав. кафедрой информационно-коммуникационных технологий, математики и информационной безопасности, директор Института заочного обучения, информационных технологий и онлайн-проектов

1 Цели и задачи изучения дисциплины (модуля).

1.1 Цель освоения дисциплины.

Цель дисциплины - формирование у магистрантов комплекса профессиональных знаний, умений и навыков в области проектирования, реализации и применения современных методов машинного обучения для решения научных и прикладных задач в сфере обработки естественного языка, цифровых коммуникаций и публичного управления.

1.2 Задачи дисциплины.

- Ознакомить магистрантов с теоретическими основами и принципами работы искусственных сетей и технологий машинного обучения, применимых в лингвистических исследованиях и цифровых коммуникациях.

- Изучить типовые технологии машинного обучения и особенности их применения в обработке естественного языка и текстовых данных.

- Сформировать практические навыки использования машинного обучения для решения задач автоматической классификации текстов, анализа тональности, извлечения информации, тематического моделирования и машинного перевода.

- Научить магистрантов применять современные технологии машинного обучения для проектирования, обучения и оценки качества нейросетевых моделей.

- Развить компетенции магистрантов в подборе оптимальных технологий машинного обучения и их интеграции в цифровые продукты и системы, используемые в цифровых коммуникациях и публичном управлении.

- Подготовить магистрантов к профессиональной деятельности, связанной с проведением исследований и разработкой прикладных решений на основе технологий машинного обучения, а также к эффективному представлению и интерпретации результатов таких решений.

1.3 Место дисциплины (модуля) в структуре образовательной программы.

Дисциплина «Б1.О.15 МАШИННОЕ ОБУЧЕНИЕ В ЛИНГВИСТИКЕ» относится к обязательной части Блока 1 «Дисциплины (модули)» учебного плана.

В соответствии с рабочим учебным планом дисциплина изучается на 1 курсе на очной форме обучения в 2 семестре. Вид промежуточной аттестации: зачет.

Для изучения дисциплины студент должен обладать знаниями по дисциплинам: «Фундаментальные основы лингвистики», «Интерпретация текста», «Математические основы и статистика для ИИ».

Полученные в процессе обучения по данной дисциплине знания могут быть использованы при изучении дисциплин «Анализ и визуализация текстовых данных», «Методы лингвистических исследований», «Медиа дискурс» и при проведении эмпирического исследования в ходе написания ВКР.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы.

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций: ОПК-6

Код и наименование индикатора*	Результаты обучения по дисциплине (знает, умеет, владеет (навыки и/или опыт деятельности))
ОПК-6	Способен осуществлять эффективное управление разработкой программных средств и информационных проектов в сфере своей профессиональной деятельности

Код и наименование индикатора*	Результаты обучения по дисциплине (<i>знает, умеет, владеет (навыки и/или опыт деятельности)</i>)
ИОПК-6.1 Владеет умениями технически проектировать программные средства и информационные проекты в сфере своей профессиональной деятельности	Знает принципы технического проектирования программных средств и информационных проектов в сфере своей профессиональной деятельности в области обработки естественного языка и текстовых данных.
	Умеет проектировать архитектуру и определять технические параметры проектирования программных средств и информационных проектов в сфере своей профессиональной деятельности в области обработки естественного языка и текстовых данных.
	Владеет навыками работы с технологиями машинного обучения в сфере своей профессиональной деятельности в области обработки естественного языка и текстовых данных.
ИОПК-6.2 Владеет способностью выявлять требования к программным средствам и информационным проектам в сфере своей профессиональной деятельности	Знает методики выявления и анализа требований к программным средствам и информационным проектам в сфере своей профессиональной деятельности.
	Умеет формулировать и конкретизировать требования к технологиям машинного обучения в области анализа текста, учитывая специфику задач и потребности целевой аудитории.
	Владеет методами и инструментами системного анализа и оценки требований к информационным проектам и нейросетевым приложениям в сфере прикладной лингвистики и цифровых коммуникаций.
ИОПК-6.3 Обеспечивает структурирование и планирование деятельности для достижения цели	Знает основы организации проектной деятельности и принципы планирования разработки и внедрения технологий машинного обучения в профессиональную деятельность.
	Умеет разрабатывать детальный план реализации проектов с использованием технологий машинного обучения, устанавливать контрольные точки и показатели успешности реализации проекта.
	Владеет технологиями планирования, управления и структурирования работы над проектами, основанными на технологиях машинного обучения, способностью оперативно корректировать планы для достижения поставленных целей.

2. Структура и содержание дисциплины.

2.1 Распределение трудоёмкости дисциплины по видам работ.

Общая трудоёмкость дисциплины составляет 3 зач.ед. (108 часов), их распределение по видам работ представлено в таблице

Для студентов ОФО.

Вид учебной работы	Всего часов	Семестр
		2
Контактная работа, в том числе:	26,2	26,2
Аудиторные занятия (всего)		

В том числе:			
Занятия лекционного типа		12	12
Лабораторные занятия		14	14
Занятия семинарского типа (семинары, практические занятия)		-	-
Иная контактная работа:		0,2	0,2
Контроль самостоятельной работы (КСР)			
Промежуточная аттестация (ИКР)		0,2	0,2
Самостоятельная работа, в том числе		81,8	81,8
Курсовая работа		-	-
Проработка учебного (теоретического) материала		21,8	21,8
Выполнение индивидуальных заданий (подготовка к лабораторным работам)		60	60
Общая трудоемкость	час.	108	108
	в том числе контактная работа	26,2	26,2
	зач. Ед	3	3

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины. Разделы (темы) дисциплины, изучаемые во 2 семестре (очная форма)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение в машинное обучение для лингвистики	14	2		2	10
2.	Модели классификации и регрессии для текстовых данных	20	2		2	16
3.	Ненаблюдаемые (Unsupervised) методы в лингвистике	20	2		2	16
4.	Инженерия признаков и методы ансамблей	22	2		4	16
5.	Оценка качества, интерпретация и практические аспекты ML в лингвистике	31,8	4		4	23,8
	<i>Итого по дисциплине:</i>	107,8	12		14	81,8
	<i>Контроль самостоятельной работы (КСР)</i>					
	<i>Промежуточная аттестация (ИКР)</i>	0,2				
	<i>Подготовка к текущему контролю</i>					
	<i>Общая трудоемкость по дисциплине</i>	108				

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов дисциплины:

2.3.1 Занятия лекционного типа.

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1.	Введение в машинное обучение для лингвистики	Введение в машинное обучение для лингвистики	Опрос по теме лекции, участие в лекции-дискуссии
2.	Модели классификации и регрессии для текстовых данных	Модели классификации и регрессии для текстовых данных	Опрос по теме лекции, участие в лекции-дискуссии
3.	Ненаблюдаемые (Unsupervised) методы в лингвистике	Ненаблюдаемые (Unsupervised) методы в лингвистике	Опрос по теме лекции, участие в лекции-дискуссии
4.	Инженерия признаков и методы ансамблей	Инженерия признаков и методы ансамблей	Опрос по теме лекции, участие в лекции-дискуссии
5.	Оценка качества, интерпретация и практические аспекты ML в лингвистике	Оценка качества, интерпретация и практические аспекты ML в лингвистике	Опрос по теме лекции, участие в лекции-дискуссии

2.3.2 Занятия лабораторного типа.

№	Наименование раздела (темы)	Тематика лабораторных занятий (семинаров)	Форма текущего контроля
1	2	3	4
1.	Введение в машинное обучение для лингвистики	Лабораторная работа № 1. Подготовка и базовый анализ текстовых данных для классификации	Презентация результатов выполнения лабораторной работы
2.	Модели классификации и регрессии для текстовых данных	Лабораторная работа № 2. Кластеризация текстов и выявление тем (topic modeling)	Презентация результатов выполнения лабораторной работы
3.	Ненаблюдаемые (Unsupervised) методы в лингвистике	Лабораторная работа № 3. Расширенные признаки и методы ансамблей	Презентация результатов выполнения лабораторной работы
4.	Инженерия признаков и методы ансамблей	Лабораторная работа № 4. Регрессионные задачи с текстовыми данными	Презентация результатов выполнения лабораторной работы
5.	Оценка качества, интерпретация и практические аспекты ML в лингвистике	Лабораторная работа № 5. Оценка и интерпретация результатов ML-моделей в лингвистике	Презентация результатов выполнения лабораторной работы

2.3.3 Примерная тематика курсовых работ (проектов)

Курсовые работы - не предусмотрены.

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

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка теоретического материала (подготовка к письменному опросу)	Методические указания по организации самостоятельной работы студентов по направлению подготовки 45.04.03 Фундаментальная и прикладная лингвистика Ученого совета факультета управления и психологии Кубанского государственного университета, протокол N 7 от 27.04.2021 г.
2	Подготовка к лабораторной работе по текущей теме	Методические указания по организации самостоятельной работы студентов по направлению подготовки 45.04.03 Фундаментальная и прикладная лингвистика Ученого совета факультета управления и психологии Кубанского государственного университета, протокол N 7 от 27.04.2021 г.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии.

В ходе изучения дисциплины предусмотрено использование следующих образовательных технологий: лабораторные занятия, проблемное обучение, модульная технология, подготовка письменных аналитических работ, самостоятельная работа студентов.

Компетентностный подход в рамках преподавания дисциплины реализуется в использовании интерактивных технологий и активных методов (проектных методик, мозгового штурма, разбора конкретных ситуаций, анализа педагогических задач, педагогического эксперимента, иных форм) в сочетании с внеаудиторной работой.

Информационные технологии, применяемые при изучении дисциплины: использование информационных ресурсов, доступных в информационно-телекоммуникационной сети Интернет.

Адаптивные образовательные технологии, применяемые при изучении дисциплины – для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Обучение в рамках дисциплины направлено на увеличение доли практической работы студента, использование игровых и имитационных форм обучения, инициирование самостоятельного поиска (студентом) знаний через проблематизацию (преподавателем) учебного материала.

В целях повышения качества профессиональной подготовки обучающихся:

- используется комплекс мультимедийных презентаций в учебном процессе;
- увеличена доля занятий, проводимых в интерактивной форме.

В преподавании курса используются современные образовательные технологии:

- информационно-коммуникативные технологии;
- исследовательские методы в обучении;
- проблемное обучение.

Для решения поставленных целей в рамках учебной дисциплины требуются использование методов обучения, направленных на формирование умений и навыков специальной аналитики. Для этого внедрены следующие образовательные технологии:

1. Проведение практического занятия, в рамках которого студенты решают двудединую задачу: а) получают знания по очередной теме учебного модуля; б) решают аналитические задачи.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. Оценочные и методические материалы

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины.

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме заданий по теме лабораторных работ, опроса в письменной форме, аналитического доклада, и других творческих заданий и контрольных работ и **промежуточной аттестации** в форме вопросов к зачету.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Код и наименование индикатора (в соответствии с п. 1.4)	Результаты обучения (в соответствии с п. 1.4)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	ИОПК-6.1 Владеет умениями технически проектировать программные средства и информационные проекты в сфере своей профессиональной деятельности	Знает принципы технического проектирования программных средств и информационных проектов в сфере своей профессиональной деятельности в области обработки естественного языка и текстовых данных. Умеет проектировать архитектуру и определять технические параметры проектирования программных средств и информационных проектов в сфере своей профессиональной деятельности в области обработки естественного языка и текстовых данных. Владеет навыками работы с технологиями машинного обучения в сфере своей профессиональной деятельности в области	Контрольная работа по теме, Лабораторная работа	Вопросы к зачету

		обработки естественного языка и текстовых данных.		
2	ИОПК-6.2 Владеет способностью выявлять требования к программным средствам и информационным проектам в сфере своей профессиональной деятельности	Знает методики выявления и анализа требований к программным средствам и информационным проектам в сфере своей профессиональной деятельности. Умеет формулировать и конкретизировать требования к технологиям машинного обучения в области анализа текста, учитывая специфику задач и потребности целевой аудитории. Владеет методами и инструментами системного анализа и оценки требований к информационным проектам и нейросетевым приложениям в сфере прикладной лингвистики и цифровых коммуникаций.	Вопросы для устного (письменного) опроса по теме, Лабораторная работа	Вопросы к зачету
3	ИОПК-6.3 Обеспечивает структурирование и планирование деятельности для достижения цели	Знает основы организации проектной деятельности и принципы планирования разработки и внедрения технологий машинного обучения в профессиональную деятельность. Умеет разрабатывать детальный план реализации проектов с использованием технологий машинного обучения, устанавливать контрольные точки и показатели успешности реализации проекта. Владеет технологиями планирования, управления и структурирования работы над проектами, основанными на технологиях машинного обучения, способностью оперативно корректировать планы для достижения поставленных целей.	Вопросы для устного (письменного) опроса по теме, Лабораторная работа	Вопросы к зачету

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Примерный перечень вопросов и заданий

Лабораторная работа 1. Подготовка и базовый анализ текстовых данных для классификации

Цель: Научиться подготавливать текстовый корпус (очистка, разметка) и применять простейшие модели классификации (Наивный Байес, Логистическая регрессия) с помощью Bag-of-Words / TF-IDF.

Основные задания:

1. Сформируйте корпус (не менее 2–3 классов, например, тематических или тональных).
2. Очистите тексты (удаление HTML-тегов, стоп-слов, возможно, лемматизация).
3. Преобразуйте тексты в векторное представление (TF-IDF, n-граммы).
4. Обучите модели (Наивный Байес, Логистическая регрессия) и сравните результаты по метрикам (accuracy, F1).
5. Проанализируйте ошибки и сделайте лингвистические выводы (какие слова или конструкции влияют на классификацию).

Результат:

- Код преобразования и классификации;

- Таблица метрик и краткий отчёт с анализом (1–2 стр. + графики / матрица ошибок).

Лабораторная работа 2. Кластеризация текстов и выявление тем (topic modeling)

Цель: Освоить методы ненаблюдаемого обучения (k-means, LDA) и проанализировать структуру больших массивов текстов.

Основные задания:

1. Подготовьте набор документов (например, новости, короткие статьи), не размеченных по классам.
2. Проведите кластеризацию (k-means или hierarchical clustering).
 - Выберите способ векторизации (TF-IDF, Doc2Vec).
 - Определите оптимальное число кластеров (элбо-метод, силуэт).
3. Реализуйте тематическое моделирование (LDA или NMF) для того же корпуса.
 - Найдите оптимальное число тем, интерпретируйте набор ключевых слов каждой темы.
4. Сравните результаты кластеризации и тематического моделирования.
 - Насколько хорошо совпадают?
 - Каковы характерные слова / топ-термины в каждом кластере / теме?

Результат:

- Визуализации (диаграммы, word cloud для каждой темы), таблица ключевых слов.
- Отчёт о том, какие лингвистические закономерности удалось обнаружить.

Лабораторная работа 3. Расширенные признаки и методы ансамблей

Цель: Научиться извлекать дополнительные лингвистические признаки и применять Random Forest или Gradient Boosting для улучшения качества классификации/регрессии.

Основные задания:

1. Возьмите корпус (тексты + метки: тема, стиль, оценка тональности и т.д.).
2. Извлеките дополнительные признаки:
 - Морфологическая информация (части речи, леммы).
 - Синтаксические зависимости (количество сложных конструкций).
 - Статистические показатели (длина предложения, уникальные слова).
3. Объедините эти признаки с TF-IDF в один вектор признаков.
4. Используйте методы ансамблей (Random Forest, XGBoost):
 - Сравните качество с простыми линейными моделями.
 - Оцените важность признаков (feature importance).
5. Сделайте выводы:
 - Какие признаки оказались самыми «сильными» для классификации?
 - Как расширенная матрица признаков повлияла на итоговое качество?

Результат:

- Код генерации признаков и обучения ансамбля.
- Сравнительные метрики (F1, точность и др.).
- Наблюдения о том, как лингвистические признаки влияют на решение модели.

Лабораторная работа 4. Регрессионные задачи с текстовыми данными

Цель: Освоить регрессию в контексте лингвистических задач (оценка сложности текста, «читаемости», предсказание рейтинга по отзывам).

Основные задания:

1. Подготовьте корпус, в котором есть числовой показатель для каждого текста (например, балл сложности, длина/глубина текста, средняя оценка пользователей).
2. Преобразуйте тексты в числовые векторы (TF-IDF или Doc2Vec).
3. Выберите и обучите регрессионную модель (линейная регрессия, Random Forest Regressor, Gradient Boosting Regressor).
4. Оцените метрики (RMSE, MAE, R^2) на тестовой выборке.
5. Проанализируйте значимость признаков или примерные интерпретации (если это позволяет модель):
 - Какие слова/структуры оказывают наибольшее влияние на предсказываемую величину?
 - Насколько точно модель может «угадать» сложность или рейтинг?

Результат:

- Код подготовки данных и обучения модели.
- Таблица метрик качества регрессии.
- Лингвистический комментарий о выявленных закономерностях.

Лабораторная работа 5. Оценка и интерпретация результатов ML-моделей в лингвистике

Цель: Научиться глубже оценивать качество моделей, понимать их ошибки и интерпретировать решения с точки зрения лингвистики.

Основные задания:

1. Возьмите любую из ранее созданных моделей (классификатор или регрессор) и выполните углублённый анализ:
 - Постройте confusion matrix (в случае классификации), разберите типичные ошибки.
 - Примените методы интерпретации (например, LIME или SHAP), чтобы увидеть, какие слова/признаки наиболее влияют на предсказание.
2. При наличии дисбаланса классов: используйте дополнительные метрики (precision, recall, F1) и методы (oversampling, undersampling).
3. Создайте отчёт об итоговом качестве:
 - Какие сегменты (жанры, подвиды текстов) лучше предсказываются, какие хуже?
 - Возможны ли улучшения и как?
4. Напишите вывод:
 - Связь полученных результатов с лингвистическими особенностями корпуса.
 - Возможные направления для дальнейшего исследования.

Результат:

- Детализированная оценка (confusion matrix, precision/recall, SHAP/LIME-графики).
- Короткое эссе (1–2 страницы) с лингвистическими и методологическими выводами.

Ключевые рекомендации по курсу

1. **Программные инструменты:**
 - Python (scikit-learn, numpy, pandas, matplotlib, NLTK, spaCy, gensim и др.), R (tidyverse, caret, tm) или иные среды.
 - Визуализация результатов (matplotlib, plotly, seaborn, библиотеки для SHAP/LIME).
2. **Корпуса:**
 - Можно использовать открытые текстовые датасеты (новостные, социальные сети, литературные тексты).

- Рекомендуется минимум 2–3 тыс. текстов для более адекватной оценки методов; при возможности – больше.
- 3. **Отчётность:**
 - Каждая лабораторная работа сопровождается отчётом (структура: краткое описание корпуса, код/псевдокод и результаты, анализ и выводы).
 - Особое внимание – лингвистической интерпретации: почему модель работает (или не работает) с тем или иным жанром, какие ошибки характерны.
- 4. **Итоговая оценка:**
 - Ориентируется на корректность реализации методов, качество анализа, полноту интерпретации результатов, а также понимание лингвистических аспектов.

Зачетно-экзаменационные материалы для промежуточной аттестации (зачет)

Вопросы к зачету.

- 1) Супервайзинг (supervised), безсупервайзинг (unsupervised), полу-супервайзинг (semi-supervised).
- 2) Примеры задач для каждой парадигмы и их лингвистические приложения.
- 3) Сбор и предварительная обработка данных (очистка, разметка).
- 4) Разделение на обучающую, валидационную и тестовую выборки.
- 5) Итеративный процесс обучения и улучшения модели.
- 6) Разнообразие жанров, наличие шума, сложность разметки (морфологической, синтаксической, семантической).
- 7) Репрезентативность корпуса, баланс классов, влияние доменных особенностей.
- 8) Bag-of-Words, TF-IDF, n-граммы.
- 9) Векторизация: основные идеи и ограничения.
- 10) Роль лемматизации и стемминга, а также фильтрация стоп-слов.
- 11) Логистическая регрессия: принцип, функции активации, оптимизация.
- 12) Наивный Байес: предпосылки, использование в анализе текстов, преимущества и недостатки.
- 13) SVM (опорные векторы): идея разделяющей гиперплоскости, примеры лингвистических задач.
- 14) В каких случаях регрессия применяется к лингвистическим данным?
- 15) Примеры числовых меток (оценка сложности, читаемости текста, предсказание рейтинга).
- 16) Метрики качества (MAE, RMSE, R^2) и их интерпретация.
- 17) Random Forest, Gradient Boosting (XGBoost, LightGBM, CatBoost).
- 18) Основные идеи бэггинга (bagging) и бустинга (boosting).
- 19) Сравнение с простыми (базовыми) методами, примеры в задачах текстовой классификации.
- 20) Различия между k-means, иерархической кластеризацией, DBSCAN.
- 21) LDA (Latent Dirichlet Allocation) и NMF (Non-negative Matrix Factorization) в тематическом моделировании.
- 22) Интерпретация и визуализация кластеров и тем (ключевые слова, распределение тем в документах).
- 23) Извлечение морфологических, синтаксических, статистических признаков из текста.
- 24) Использование POS-тегов, коллокаций, частот синтаксических конструкций.
- 25) Как расширенные признаки улучшают работу модели.
- 26) Accuracy, precision, recall, F1, AUC (ROC-кривая) и когда их стоит применять.
- 27) Работа с несбалансированными классами (метрики macro/micro-averaging и др.).

- 28) Интерпретация confusion matrix, наиболее частые ошибки в лингвистическом ML.
- 29) Применение кластеризации к текстам (группировка по темам/жанрам).
- 30) Преимущества и недостатки ненаблюдаемых методов для лингвистических задач.
- 31) Снижение размерности (PCA, t-SNE, UMAP) и их лингвистические интерпретации.
- 32) Подходы LIME и SHAP, визуализация важности признаков (feature importance).
- 33) Какие слова или конструкции чаще всего влияют на решение модели?
- 34) Значимость интерпретации для лингвистических исследований и практических приложений.
- 35) Методы борьбы: регуляризация (L1/L2), ранняя остановка (early stopping), кросс-валидация и др.
- 36) Специфика переобучения в задачах анализа естественного языка.
- 37) Работа с большими корпусами: параллелизация, распределённые вычисления.
- 38) Инструменты и фреймворки (scikit-learn, spaCy, NLTK, gensim).
- 39) Лучшие практики документирования экспериментов и управления версиями данных (MLflow и др.).
- 40) В каких случаях оправданы классические модели?
- 41) Плюсы и минусы глубокого обучения (векторные представления, требования к данным, вычислительные ресурсы).
- 42) Примеры гибридных решений (комбинация классического ML с нейросетевыми эмбедами).

Перечень части компетенции, проверяемых оценочным средством:

- ИОПК-6.1 Владеет умениями технически проектировать программные средства и информационные проекты в сфере своей профессиональной деятельности
- ИОПК-6.2 Владеет способностью выявлять требования к программным средствам и информационным проектам в сфере своей профессиональной деятельности
- ИОПК-6.3 Обеспечивает структурирование и планирование деятельности для достижения цели

Критерии оценки:

«зачтено» - свободное владение практическим материалом в рамках учебной дисциплины, полные развернутые ответы на вопросы на зачете, умение формализовать практическую задачу по профилю своей специальности и решить её с использованием изученных особенностей работы с информацией, подготовка всех практических заданий,

«не зачтено» - недостаточное владение практическим материалом, отсутствие навыков использования информационных технологий для решения практических задач по профилю своей специальности, не выполнение лабораторных работ и итогового проекта.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля).

5.1 Учебная литература:

1. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662> (дата обращения: 15.02.2025).

2. Нугуманова, А. Б. Автоматизированная обработка текстовых массивов : учебник и практикум для вузов / А. Б. Нугуманова. — 2-е изд. — Москва : Издательство Юрайт, 2024. — 82 с. — (Высшее образование). — ISBN 978-5-534-20738-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558668> (дата обращения: 15.02.2025).

3. Советов, Б. Я. Информационные технологии : учебник для вузов / Б. Я. Советов, В. В. Цехановский. — 8-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 414 с. — (Высшее образование). — ISBN 978-5-534-20054-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559897> (дата обращения: 15.02.2025).

5.2. Периодическая литература

1. Вестник Волгоградского государственного университета. Серия 2: Языкознание
2. Вестник Московского университета. Серия 19: Лингвистика и межкультурная коммуникация
3. Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация
4. Вестник Южно-Уральского государственного университета. Серия: Лингвистика
5. Вестник Южно-Уральского государственного университета. Серия: Социально-гуманитарные науки
6. Медиалингвистика
7. Медиаскоп
8. Политическая лингвистика.
9. Социалингвистика
10. Теоретическая и прикладная лингвистика

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

1. Электронный каталог
2. Поступления литературы в библиотеки филиалов
3. Поступления диссертаций и авторефератов
4. Статьи из периодики и научных сборников с 2016 г.
5. Статьи из периодики и научных сборников до 2016г.
6. Газеты и журналы
7. Электронная библиотека трудов ученых КубГУ

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» www.biblioclub.ru
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>
6. ЭБ ОИЦ «Академия» <https://academia-moscow.ru/elibrary/>

Профессиональные базы данных (российские):

1. Виртуальный читальный зал Российской государственной библиотеки (РГБ) <https://ldiss.rsl.ru/>
2. Национальная электронная библиотека <https://rusneb.ru/>
3. Базы данных компании «ИВИС» <https://eivis.ru/>
4. Научная электронная библиотека eLIBRARY.RU (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовая коллекция журналов на платформе РЦНИ (Электронные версии научных журналов РАН) <https://journals.rcsi.science/>
6. Президентская библиотека им. Б.Н. Ельцина <https://www.prlib.ru/>
7. Электронная библиотечная система социо-гуманитарного знания «SOCHUM» <https://sochum.ru/>

Профессиональные базы данных (зарубежные):

1. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
2. Полнотекстовая коллекция книг eBook Collections издательства SAGE Publications <https://sk.sagepub.com/books/discipline>
3. Полнотекстовая коллекция книг EBSCO eBook <https://books.kubsu.ru/>
4. Ресурсы Springer Nature <https://link.springer.com/>, <https://www.nature.com/>
5. Questel. База данных Orbit Premium edition <https://www.orbit.com>
6. Полнотекстовые коллекции книг издательства American Institute of Physics Publishing (AIPP Ebook) <https://pubs.aip.org/books>
7. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
8. China National Knowledge Infrastructure. БД CNKI Academic Reference (AR) <https://ar.oversea.cnki.net/>

Информационные справочные системы:

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Базы данных открытого доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Лекториум ТВ - видеолекции ведущих лекторов России <http://www.lektorium.tv/>

Ресурсы свободного доступа:

1. КиберЛенинка (<http://cyberleninka.ru/>);
2. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
3. Федеральный портал "Российское образование" <http://www.edu.ru/>;
4. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
5. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/>
6. Федеральный центр информационно-образовательных ресурсов (<http://fcior.edu.ru/>);
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;
11. Образовательный портал "Учеба" <http://www.ucheba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.xn--plai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы

КубГУ:

1. Открытая среда модульного динамического обучения КубГУ <https://openedu.kubsu.ru/>
2. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
3. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины (модуля).

Лекционные и лабораторные занятия – являются формой учебной аудиторной работы, в рамках которой формируются, закрепляются и представляются студентами знания, умения и навыки, интегрирующие результаты освоения компетенций как в лекционном формате, так в различных формах самостоятельной работы. К каждому занятию преподавателем формулируются практические задания, требования и методические рекомендации к их выполнению, которые представляются в фонде оценочных средств учебной дисциплины.

Контроль самостоятельной работы: для студентов дневной формы обучения – текущий контроль осуществляется в соответствии с программой занятий (еженедельно для студентов очной формы обучения); промежуточный контроль по итогам освоения дисциплины осуществляется в форме рейтинговой системы оценок. Описание заданий для самостоятельной работы студентов и требований по их выполнению выдаются преподавателем в соответствии с разработанным фондом оценочных средств.

Самостоятельная работа студентов по данному учебному курсу предполагает поэтапную подготовку по каждому разделу в рамках соответствующих заданий.

Первый этап самостоятельной работы студентов включает в себя тщательное изучение теоретического материала, рекомендуемых разделов основной и дополнительной

литературы, материалов периодических научных изданий, необходимых для овладения понятийно-категориальным аппаратом и формирования представлений о комплексе аналитического инструментария, используемого в рамках данной отрасли знания;

На втором этапе на основе сформированных знаний и представлений по данному разделу студенты выполняют лабораторные задания, нацеленные на формирование умений и навыков в рамках заявленной компетенции. На данном этапе студенты осуществляют самостоятельный поиск эмпирических материалов в рамках конкретного задания, обобщают и анализируют собранный материал по схеме, рекомендованной преподавателем, формулируют выводы, готовят практические рекомендации, презентационные материалы для публичного их представления и обсуждения.

Критерии оценки заданий в рамках самостоятельной работы студентов формулируются преподавателем в фонде оценочных средств.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Материально-техническое обеспечение по дисциплине (модулю)

По всем видам учебной деятельности в рамках дисциплины используются аудитории, кабинеты и лаборатории, оснащенные необходимым специализированным и лабораторным оборудованием%

- использование электронной почты для общения со студентами в рамках учебного курса;

- методы обучения с использованием информационных технологий (демонстрация мультимедийных материалов);

- технические средства: компьютерная техника (ноутбук, проектор, экран).

Наименование специальных помещений	Оснащенность специальных помещений	Перечень лицензионного программного обеспечения
Учебные аудитории для проведения лабораторных работ и лекционных занятий: ауд. 324	Мебель: учебная мебель Технические средства обучения: экран, проектор, телевизор, компьютеры, электронные ресурсы, доска учебная. Оборудование: стационарный компьютер с доступом в Интернет, доступ в электронную информационно-образовательную среду.	Для подготовки и демонстрации презентационных материалов используется пакет PowerPoint Microsoft Office, ОС Microsoft Windows 10 выходом в Интернет. Для проведения лабораторных работ используется программное обеспечение Gephy.
Компьютерный класс, Аудитория курсового проектирования, Научно-учебная лаборатория "Лингвистика и кросс-культурная коммуникация", Аудитория для текущего	Мебель: учебная мебель Технические средства обучения: экран, проектор, телевизор, компьютеры, электронные ресурсы, доска учебная. Оборудование: стационарный компьютер с	Для подготовки и демонстрации презентационных материалов используется пакет PowerPoint Microsoft Office, ОС Microsoft Windows 10 выходом в Интернет. Для проведения

контроля и промежуточной аттестации: ауд. 320	доступом в Интернет, доступ в электронную информационно-образовательную среду.	лабораторных работ используется программное обеспечение Gerphy.
Учебная аудитория для проведения занятий семинарского типа, текущего контроля и промежуточной аттестации: ауд. 364	Мебель: учебная мебель Технические средства обучения: экран, проектор, телевизор, компьютеры, электронные ресурсы, доска учебная. Оборудование: стационарный компьютер с доступом в Интернет, доступ в электронную информационно-образовательную среду.	Для подготовки и демонстрации презентационных материалов используется пакет PowerPoint Microsoft Office, ОС Microsoft Windows 10 выходом в Интернет. Для проведения лабораторных работ используется программное обеспечение Gerphy.

Для самостоятельной работы обучающихся предусмотрены помещения, укомплектованные специализированной мебелью, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование помещений для самостоятельной работы обучающихся	Оснащенность помещений для самостоятельной работы обучающихся	Перечень лицензионного программного обеспечения
Помещение для самостоятельной работы обучающихся (читальный зал Научной библиотеки)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	Для подготовки и демонстрации презентационных материалов используется пакет PowerPoint Microsoft Office, ОС Microsoft Windows 10 выходом в Интернет.

Рецензия
на рабочую программу дисциплины
Б1.О.15 Машинное обучение в лингвистике
направления 45.04.03 Фундаментальная и прикладная лингвистика
направленность Прикладные языковые технологии в цифровых
коммуникациях и публичном управлении,
разработанная на кафедре прикладной лингвистики и новых
информационных технологий
ФГБОУ ВО «Кубанский государственный университет»

Рабочая программа дисциплины «Б1.О.15 Машинное обучение в лингвистике» разработана на основе ФГОС ВО 45.04.03 Фундаментальная и прикладная лингвистика (Приказ Минобрнауки России от 12.08.2020 № 993), отражает его требования и соответствует специфике подготовки лингвистов по заявленной направленности Прикладные языковые технологии в цифровых коммуникациях и публичном управлении.

Разделы рабочей программы детализировано описывают цели и задачи изучения дисциплины, планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы. Содержание дисциплины включает распределение видов учебной работы и их трудоемкости по разделам дисциплины. Также в программе представлен перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю) и структура оценочных средств для текущей и промежуточной аттестации, содержащая в том числе типовые контрольные задания, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы. Перечень литературы составлен в соответствии с требованиями ФГОС.

Рецензируемая рабочая программа отвечает требованиям, предъявляемым организации учебного процесса по направлению 45.04.03 Фундаментальная и прикладная лингвистика и отражает требования направленности магистерской программы. Рекомендую представленную рабочую программу дисциплины для организации обучения в рамках программы магистратуры «Прикладные языковые технологии в цифровых коммуникациях и публичном управлении».

Директор Института заочного обучения, информационных технологий и онлайн-проектов, зав. кафедрой информационно-коммуникационных технологий, математики и информационной безопасности, профессор, доктор философских наук



Г.А. Воробьев

Рецензия
на рабочую программу дисциплины
«Б1.О.15 Машинное обучение в лингвистике»
направления 45.04.03 Фундаментальная и прикладная лингвистика,
направленность Прикладные языковые технологии в цифровых
коммуникациях и публичном управлении,
разработанная на кафедре социальной работы, психологии и педагогики
высшего образования
ФГБОУ ВО «Кубанский государственный университет»

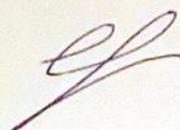
Рабочая программа дисциплины «Б1.О.15 Машинное обучение в лингвистике» относится к обязательной части Блока 1 «Дисциплины (модули)» учебного плана и ориентирована на формирование у магистрантов комплекса профессиональных знаний, умений и навыков в области проектирования, реализации и применения современных методов машинного обучения для решения научных и прикладных задач в сфере обработки естественного языка, цифровых коммуникаций и публичного управления.

Рабочая программа дисциплины включает все необходимые разделы: цель и задачи дисциплины; место дисциплины в структуре основной образовательной программы и её взаимосвязь с другими дисциплинами; компетенции, формируемые в процессе обучения и требования к результатам обучения, объем дисциплины в соответствии с видами учебной работы в часах, структуру курса по темам, содержание дисциплины; формы текущей и промежуточной аттестации; учебно-методическое и информационное обеспечение дисциплины.

Автор в своей программе предусмотрела разнообразные интерактивные образовательные технологии, соответствующие компетентностному подходу и используемые на более, чем $\frac{1}{2}$ части занятий в процессе обучения.

Таким образом, рецензируемая рабочая программа дисциплины составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования, отвечает требованиям, предъявляемым организации учебного процесса по направлению 45.04.03 Фундаментальная и прикладная лингвистика и отражает требования направленности магистерской программы по соответствующему направлению подготовки и может быть использована в учебном процессе ФГБОУ ВО «КубГУ».

Канд. физ.-мат. наук, доцент кафедры общего,
стратегического, информационного менеджмента
и бизнес-процессов, КубГУ



А.П. Савченко