



Основная профессиональная образовательная программа
45.04.01 Филология (Прикладная цифровая филология)

Министерство науки и высшего образования Российской Федерации

ИВАНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Кафедра зарубежной филологии

ОДОБРЕНО:

Руководитель ОП

_____ С.А. Маник

(подпись)

«__29__» _____ сентября ____2024__ г.

Рабочая программа дисциплины

Машинное обучение

Уровень высшего образования:	магистратура
Квалификация выпускника:	магистр
Направление подготовки:	45.04.01 Филология
Направленность (профиль) образовательной программы:	Прикладная цифровая филология

Иваново 2024



Основная профессиональная образовательная программа

45.04.01 Филология (Прикладная цифровая филология)

1. Цели освоения дисциплины

Целью освоения дисциплины "Машинное обучение" является обучение магистрантов навыкам использования машинного обучения и управления данными для последующей разработки стратегий продвижения лингвистических продуктов и услуг в цифровом пространстве, а также создания цифрового продукта в своей профессиональной деятельности.

Задачи дисциплины:

- изучение основ разработки учебных курсов в среде Moodle для дистанционного обучения;
- приобретение навыков работы на языке Python, знание и понимание задач управления данными, в том числе, загрузка данных, преобразование данных, и предварительный анализ, и визуализация данных;
- знакомство с основными задачами и моделями машинного обучения, знание методов оценки качества работы различных моделей машинного обучения;
- понимание процесса интеграции моделей машинного обучения в рамках задач, стоящих перед цифровыми филологами;
- повышение мотивации студентов на дальнейшее углубление своих знаний в области работы с данными и машинного обучения.

Курс ориентирован на магистров ОП «Прикладная цифровая филология», примеры и задачи подбираются с учетом профессиональных интересов магистров. Особое внимание уделяется практическому освоению терминологии и определениям базовых терминов.

2. Место дисциплины в структуре ОП

Дисциплина относится к факультативной части образовательной программы, формируемой участниками образовательных отношений ФТД.01.

Студент, приступающий к изучению дисциплины, должен обладать знаниями, умениями, навыками/опытом практической деятельности, полученными ранее в ходе изучения дисциплин на уровне бакалавриата. Для освоения данной дисциплины обучающийся должен владеть следующими знаниями и компетенциями:

- Основы программирования
- Компьютерная лингвистика
- Статистика

Успешное освоение данной дисциплины способствует готовности студентов к освоению дисциплины Основы иностранных языков: теория и практика перевода, Применение Big Data в гуманитарной сфере, а также опыта профессиональной деятельности производственной практики, научно-исследовательской работы и производственной практики и предваряет научно-исследовательскую работу по направлению «Филология», направленность «Прикладная цифровая филология».

Данная учебная дисциплина опирается на материал предшествующих теоретических курсов «Филология в системе гуманитарного знания», «Информационные технологии в науке и образовании», «Основы программирования для филологов», «Управление цифровыми проектами», «Терминоведение и базы данных», «Цифровые инструменты в переводе».

3. Планируемые результаты обучения по дисциплине

3.1. Компетенции, формированию которых способствует дисциплина

При освоении дисциплины формируются следующие компетенции в соответствии с ФГОС ВО по данному направлению подготовки:

ПК 4 - Способен самостоятельно осуществлять полный цикл создания, обработки, редактирования и продвижения цифрового контента разных жанров, дискурсов



Основная профессиональная образовательная программа 45.04.01 Филология (Прикладная цифровая филология)

ПК-5 - Способен использовать современные информационные технологии, управлять данными с использованием прикладных программ, баз данных и корпусов текстов, выполнять перевод текстов на иностранный язык с их помощью

3.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения формируемых компетенций

В результате освоения дисциплины обучающийся должен:

Знать:

- теоретические, методологические и практические основы использования компьютерных технологий в лингвистике, мета язык компьютерной и корпусной лингвистики (ПК-4, ПК-5);
- основные методы машинного обучения и условия их применимости (ПК-4, ПК-5)
- общие принципы устройства и работы современной вычислительной техники (ПК-5);
- подходы и этапы обработки текста на естественного языка (ПК-4, ПК-5);
- алгоритмы и этапы компьютерного (автоматического) анализа текста (ПК-4, ПК-5);

Уметь:

- использовать готовые библиотеки машинного обучения (ПК-1);
- разрабатывать программы, реализующие алгоритмы машинного обучения, на языке Python (ПК-4, ПК-5)
- использовать возможности и инструменты ИИ и компьютерных технологий при решении собственных задач в исследовательской и практической работе (ПК-4, ПК-5);
- применять полученные лингвистические знания и изученные вычислительные методы и алгоритмы к реальным лингвистическим данным (ПК-5);
- самостоятельно изучать, подключать и использовать компьютерные программы и модули, необходимые для решения конкретных задач компьютерной лингвистики (ПК-4, ПК-5);

Иметь практический опыт/Иметь навыки:

- использовать инструменты для работы с текстами естественного языка, (ПК-4; ПК-5);
- решать прикладные задачи с помощью машинного обучения (ПК-4. ПК-5);
- внедрять наиболее популярные компьютерные лингвистические инструменты, библиотеки и приложения для работы с лингвистическими данными (ПК-4, ПК-5).

4. Объем и содержание дисциплины

Объем дисциплины составляет 1 зачетных единиц (36 академических часов), в т.ч.: практическая подготовка (ПП) – 14 академических часов в очной форме

4.1. Содержание дисциплины по разделам (темам), соотнесенное с видами и трудоемкостью занятий лекционно-семинарского типа

Объем иной контактной работы и самостоятельной работы обучающегося по дисциплине указан в учебном плане образовательной программы.

№ п/п	Разделы (темы) дисциплины	Семестр	Виды занятий, их объем (в ак. часах, по очной форме обучения)		Формы текущего контроля успеваемости (по очной форме обучения)
			Занятия лекционного типа	Занятия семинарского типа	Формы промежуточной аттестации



Основная профессиональная образовательная программа
45.04.01 Филология (Прикладная цифровая филология)

1.	Вводный. Понятие больших данных и управление данными. Введение в машинное обучение	3		2 практ. занятие	Входная диагностика: тест с последующим обсуждением результатов. Список вопросов, интересующих студента по содержанию дисциплины (сдается в письменном виде)
2.	Практика работы с данными на основе задачи кластеризации	3		2 практ. занятие	Решение учебных задач
3.	Практическое построение моделей на основе логистической регрессии	3		2 практ. занятие	Решение учебных задач
4.	Метрики качества в машинном обучении	3		2 практ. занятие	Решение учебных задач
5.	Ансамбли моделей в машинном обучении	3		2 практ. занятие	Решение учебных задач
6.	Нейронные сети: сверточные и рекуррентные сети	3		2 практ. занятие	Решение учебных задач
7.	Заключительный. Подведение и анализ промежуточных результатов освоения дисциплины	3		2 практ. занятие	Решение учебных задач
Итого за семестр:				14	Зачет

4.2. Развернутое описание содержания дисциплины по разделам (темам)

Раздел 1. Вводный. Понятие больших данных и управление данными. Введение в машинное обучение

Содержание раздела:

- история появления науки о данных;
- определения и термины;
- приложения и системы в больших данных;
- как началась эра больших данных;
- откуда берутся большие данные;
- машинные данные и способы их получения;
- интеграция разнородных данных;
- типология задач машинного обучения;
- обучение с учителем, обучение без учителя, частичное обучение, обучение с подкреплением
- Визуализация данных

Раздел 2. Практика работы с данными на основе задачи кластеризации

- метод кластеризация K-средних
- другие методы кластеризации
- методы понижения размерности (метод главных компонент)

Раздел 3. Практическое построение моделей на основе логистической регрессии

- математическая основа метода логистической регрессии
- работа с признаками модели
- проблема переобучения
- использование регуляризаций L1 и L2
- оценка качества модели



Раздел 4. Метрики качества в машинном обучении

- базовые метрики качества, используемые в задачах классификации
- базовые метрики качества, используемые в задачах регрессии
- базовые метрики качества, используемые в задачах кластеризации

Раздел 5. Ансамбли моделей в машинном обучении

- основная мотивация в использовании ансамблей моделей;
- Random Forest
- Gradient Boosting

Раздел 6. Нейронные сети: сверточные и рекуррентные сети

- Базовые понятия нейронных сетей
- принцип обратного распространения ошибки
- машины Больцмана
- автоэнкодеры, GAN
- обзор базовых элементов сверточных нейронных сетей
- обзор базовых элементов рекуррентных нейронных сетей

Раздел 7. Заключительный. Подведение итогов и анализ промежуточных результатов освоения дисциплины

5. Образовательные технологии

Рекомендуемые образовательные технологии: практические занятия с технологиями учебной дискуссии, проектная работа, представляемая в виде презентаций, технологии смешанного обучения, индивидуальные задания, самостоятельная работа студентов, контрольные работы. Студенты знакомятся с компьютерными и корпусными технологиями; применением машинного обучения в типичных случаях практики, разбирают практические кейсы, участвуют в лекциях приглашенных докладчиков.

При проведении практических занятий рекомендуется помимо обсуждения теоретических вопросов использование проектных методик, выполнение практических заданий на основе Интернет-ресурсов. Студенты представляют результаты своей самостоятельной и практической работы в виде презентаций с использованием программы PowerPoint и публикации материалов в ЭИОС

6. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся организована в виде изучения теоретического материала, решения задач и практического применения полученных навыков для машинного обучения и управления большими данными.

Студентам на ресурсе «Мой университет» выкладываются ссылки на записанные вебинары по разным темам, которые они могут посмотреть в удобное время; список площадок дистанционных курсов, на которые они должны подписаться и прислать отчет прохождения (скриншоты по неделям и успеваемость по курсу).

Полностью весь методический материал по обеспечению самостоятельной работы студентов приводится в Приложении 1 к РП.

7. Характеристика оценочных средств для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Вопросы к обсуждению в ходе собеседования. Индивидуальные задания/работы с постановкой конкретной задачи и предписанным алгоритмом ее выполнения. Подготовка проектов.

Примерные вопросы к устному опросу.

Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.

Введение в анализ данных и язык Python.

Введение в анализ данных.

Введение в язык Python.

Основные операции работы с данными.

Визуализация данных.

Фундаментальные основы машинного обучения.

Статистика и машинное обучение.

Задачи классификации и регрессии.

Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.

Метрики качества алгоритмов регрессии и классификации.

Линейная регрессия. Простая многомерная регрессия. Регрессия с полиномиальными признаками. Методы регуляризации: Ridge, Lasso, ElasticNet.

Логистическая регрессия.

Деревья решений. Методы построения деревьев. Их регуляризация.

Случайный лес, его особенности.

Композиции алгоритмов. Разложение ошибки на смещение и разброс.

Бэггинг и бустинг. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.

Задача кластеризации.

Прикладной опыт ведения проектов машинного обучения.

Разработка и создание аналитических приложений.

Внедрение, сопровождение и мониторинг модели.

Экспертиза качества моделей.

Критерии оценки опроса:

Оценка	Критерии
Отлично	Отвечает на все вопросы, продемонстрирована способность анализировать и обобщать информацию, сделаны обоснованные выводы, разъяснения
Хорошо	Отвечает не на все вопросы, но на большинство. Умеет в некоторой степени анализировать и обобщать информацию. Выводы умеет делать
Удовлетворительно	Отвечает не на все вопросы (меньше 50%). Демонстрирует некоторые способности анализировать и обобщать информацию. Выводы делать не может
Неудовлетворительно	Задание не выполнено

Также студенты выполняют индивидуальные задания по каждой теме, представляя результаты анализа и синтеза данных.



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

В конце магистранты сдают зачет. Если студент присутствовал на 60% занятий и выполнил не менее 5 заданий на практических заданиях, то он получает «зачет» автоматически как накопительную оценку. Если данное требование не выполнено, то проходит собеседование с преподавателем.

Типовые варианты тестовой работы представлены в фонде оценочных средств (Приложение 2).

8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература:

1. Моисеева, И. Ю. Квантитативная лингвистика и новые информационные технологии : учебное пособие / И. Ю. Моисеева ; Оренбургский государственный университет. – Оренбург : Оренбургский государственный университет, 2017. – 103 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=481797> (дата обращения: 06.09.2024). – Библиогр. в кн. – ISBN 978-5-7410-1713-5. – Текст : электронный.
2. Компаниец, В. С. Проектирование и юзабилити-исследование пользовательских интерфейсов : учебное пособие : [16+] / В. С. Компаниец, А. Е. Лызь ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2020. – 107 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=619064>.
3. Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие : [16+] / Н. А. Бутаков, М. В. Петров, Д. Насонов. – Санкт-Петербург : Университет ИТМО, 2019. 52 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=566771>.
4. Костюк, А. И. Организация облачных и GRID-вычислений : учебное пособие : [16+] / А. И. Костюк. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2018. – 122 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=561079>.

Дополнительная литература:

1. [Электронный ресурс biblioclub: <http://biblioclub.ru/index.php?page=book&id=362986>] Моделирование систем: Подходы и методы: учебное пособие / В.Н. Волкова, Г.В. Горелова, В.Н. Козлов и др.; Министерство образования и науки Российской Федерации, Санкт-Петербургский государственный политехнический университет. – Санкт-Петербург: Издательство Политехнического университета, 2013. – 568 с.
2. Ляшевская, О. Н. Корпусные инструменты в грамматических исследованиях русского языка / О. Н. Ляшевская. – Москва : Языки славянской культуры (ЯСК) : Рукописные памятники Древней Руси, 2016. – 520 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=473302> (дата обращения: 06.09.2024). – Библиогр.: с. 480-513. – ISBN 978-5-9907947-8-8. – Текст : электронный.
3. [Электронный ресурс biblioclub: <http://biblioclub.ru/index.php?page=book&id=480500>] Жуковский О. И. Информационные технологии и анализ данных: учебное пособие / О.И. Жуковский - Томск: Эль Контент, 2014. - 130 с.
4. Электронный ресурс: <http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf> Harald Baayen (2008). Analyzing Linguistic Data. A Practical Introduction to Linguistics using R, Cambridge University Press: Cambridge

Ресурсы информационно-телекоммуникационной сети «Интернет»:



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Система электронной поддержки образовательного процесса «Мой университет» <https://uni.ivanovo.ac.ru>

Профессиональные базы данных и информационно-справочные системы:

ЭБС «Университетская библиотека онлайн» www.biblioclub.ru;

<http://lib.ivanovo.ac.ru/index.php/polnotekstovye-resursy/ebs-universitetskaya-biblioteka>

Электронная библиотека ИвГУ <http://lib.ivanovo.ac.ru/index.php/polnotekstovye-resursy/elibnew>

Электронный каталог НБ ИвГУ <http://lib.ivanovo.ac.ru/index.php/ek>

Программное обеспечение: операционная система Microsoft Windows, пакет офисных программ Microsoft Office и(или) LibreOffice, интернет-браузер Microsoft Edge и(или) Yandex Browser.

9. Материально-техническое обеспечение дисциплины

Учебные аудитории:

- для проведения занятий семинарского типа, консультаций, текущего контроля и промежуточной аттестации с комплектом специализированной учебной мебели и техническими средствами обучения;

Лаборатория, оснащенная лабораторным оборудованием, комплектом специализированной учебной мебели и техническими средствами обучения.

Помещение для самостоятельной работы, оснащенное комплектом специализированной учебной мебели, компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в ЭИОС.

Демонстрационное оборудование и учебно-наглядные пособия для занятий лекционного типа, обеспечивающие тематические иллюстрации: презентации PowerPoint, электронные словари и справочно-информационные порталы, видеоматериалы и т.п.

Автор(ы) рабочей программы дисциплины: доцент кафедры английской филологии, д.ф.н., доц. С.А. Маник

Программа рассмотрена и утверждена на заседании кафедры зарубежной филологии

«29» августа 2024 г., протокол № 13.

Программа обновлена

протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)

Программа обновлена

протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Программа обновлена
протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)