

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям


О.Н. Здрок
«30» июня 2020 г.

Регистрационный № УД 8449/уч.



ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

**Учебная программа учреждения высшего образования по учебной
дисциплине для специальности**

1-31 80 03 Математика и компьютерные науки

профилизация

Веб-программирование и интернет-технологии

Математическое и программное обеспечение мобильных устройств

2020 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 80 03-2019 и учебных планов G31-031/уч., G31-033/уч., G31з-032/уч., G31з-034/уч. от 11.04.2019

СОСТАВИТЕЛЬ:

Гуревский Алексей Николаевич, старший преподаватель кафедры веб-технологий и компьютерного моделирования Белорусского государственного университета.

РЕЦЕНЗЕНТ:

Ковалева Ирина Львовна, кандидат технических наук, доцент кафедры программного обеспечения информационных систем и технологий Белорусского национального технического университета.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой Веб-технологий и компьютерного моделирования
(протокол № 9 от 20.05.2020 г.);

Научно-методическим советом Белорусского государственного университета
(протокол № 5 от 17.06.2020 г.)

Заведующий кафедрой _____ В.М. Волков

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Целью дисциплины «Интеллектуальный анализ данных» является ознакомление студентов магистратуры с основными современными способами автоматического обнаружения и описания закономерностей в данных из окружающего мира, обучение способам автоматизации деятельности на основе анализа данных и моделирования, без явного предварительного написания алгоритмов действий, подготовка специалистов, способных использовать фундаментальные математические знания в качестве основы при проведении прикладных исследований.

Задачи учебной дисциплины:

- формирование у магистрантов способностей самостоятельно разрабатывать алгоритмы решения задач и их анализировать;
- развивать и использовать инструментальные средства, информационные среды, автоматизированные системы;
- использовать математические и компьютерные методы исследований при анализе современных естественнонаучных, экономических, социально-политических процессов;
- приобретение способностей самостоятельно расширять компьютерные математические знания с дальнейшим их использованием при анализе математических моделей широкого круга прикладных задач.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием (магистра): учебная дисциплина имеет прикладную направленность и готовит магистрантов к освоению и реализации методов интеллектуального анализа данных.

Учебная дисциплина «Интеллектуальный анализ данных» относится к модулю «Анализ данных и методы искусственного интеллекта» компонента учреждения высшего образования и адресована студентам 2-го года обучения в магистратуре по специальности 1-31 80 03 Математика и компьютерные науки, профилизация «Веб-программирование и интернет-технологии» и «Математическое и программное обеспечение мобильных устройств».

Изучение дисциплины «Интеллектуальный анализ данных» базируется на знаниях, полученных при изучении дисциплины «Машинное обучение».

Требования к компетенциям

Освоение учебной дисциплины «Интеллектуальный анализ данных» должно обеспечить формирование следующих *специализированных компетенций*:

СК-2. Быть способным эффективно использовать инструменты анализа и алгоритмы обработки данных.

В результате изучения дисциплины студент магистратуры должен:
знать:

- классификацию видов данных, имеющихся в окружающем мире;

- формулировку задач регрессии, классификации, ранжирования, оценки распределения, уменьшения размерности, кластеризации;

- основные подходы к решению задач вышеупомянутых задач регрессии, классификации, ранжирования, оценки распределения, уменьшения размерности, кластеризации;

уметь:

- определять вид имеющихся данных;

- проводить первичную обработку данных;

- формализовывать задачу, как задачу регрессии, классификации, ранжирования, оценки распределения, уменьшения размерности, кластеризации или их совокупность;

- решать упомянутые задачи с помощью современных программных средств;

- оценивать качество предложенного решения;

владеть:

- способами автоматизации деятельности на основе анализа данных и моделирования, без явного предварительного написания алгоритмов действий.

Структура учебной дисциплины

В соответствии с учебными планами специальности на изучение дисциплины отводится:

Профилизация, форма обучения	Срок обучения, лет	Курс	Семестр	Зачетные единицы	Экзамен, семестр	Зачет, семестр	Всего часов	В том числе ауд.	Из них	
									лекций	лабораторных занятий
Веб-программирование и интернет технологии, дневная	1.8	2	3	6		3	198	72	36	36
Математическое и программное обеспечение мобильных устройств, дневная	1.8	2	3	6		3	198	72	36	36
Веб-программирование и интернет технологии, заочная	2	2	3	6		3	198	16	8	8
Математическое и программное обеспечение мобильных устройств, заочная	2	2	3	6		3	198	16	8	8

Трудоемкость учебной дисциплины составляет 6 зачетных единиц.

Форма текущей аттестации по учебной дисциплине – дневная и заочная формы обучения – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Предмет Data Mining. Прикладные задачи. Этапы решения. Модели. Методы предварительной подготовки данных. Получение данных из разных источников. Проверка качества данных. Визуализация данных. Метрики оценки качества моделей. Accuracy, precision и recall. Logistic Loss.

Тема 2. Снижение размерности. Эффект проклятия размерностей. Отбор и выделение признаков. Факторный анализ. Метод главных компонент.

Тема 3. Обучение с «учителем». Задачи классификации. Сравнение классификаторов.

Тема 4. Обучение без «учителя». Задачи кластеризации. K-среднее. Ближайшие соседи.

Тема 5. Модели: иерархические, вероятностные, составные. Дерево решений. Байесовская сеть. Марковская сеть. Bagging. Boosting. Random forest.

Тема 6. Прогнозирование временных рядов. Примеры временных рядов. Стационарность в узком и широком смысле. Модели: AR(p), MA(q), ARMA(p, q), ARIMA(p,n,q), GARCH.

Тема 7. Нейронные сети. Сбор данных. Пред-/постобработка данных. Многослойный перцептрон. Глубокое обучение. Сверточные сети.

Тема 8. Обработка естественного текста (NLP). Поиск предложений, определение частей речи. Классификация текстов и документов. Выделение взаимосвязей.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

дневная форма получения образования

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УРС	Литература	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное			
1	Предмет Data Mining	2			2			[5] о	Опрос, защита лабораторной работы
2	Снижение размерности	4			4			[1, 2] о	Опрос, защита лабораторной работы
3	Обучение с «учителем»	4			4			[5] о	Опрос, защита лабораторной работы
4	Обучение без «учителя»	4			4			[5] о	Опрос, защита лабораторной работы
5	Модели: иерархические, вероятностные, составные	4			4			[2, 5] о, [5] д	Опрос, защита лабораторной работы
6	Прогнозирование временных рядов	6			6			[1, 2] о [1, 3] д	Опрос, защита лабораторной работы, творческое задание
7	Нейронные сети	6			6			[1-2] о, [5] д	Опрос, защита лабораторной работы, творческое задание

8	Обработка естественного текста (NLP)	6			6			[3] о, [2] д	Опрос, защита лабораторной работы
	ВСЕГО	36			36				Зачет

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

заочная форма получения образования

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УРС	Литература	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное			
1 2 3 4	Предмет Data Mining. Снижение размерности. Обучение с «учителем». Обучение без «учителя»	2			2			[1, 2, 5] о	Опрос, защита лабораторной работы
5 6	Модели: иерархические, вероятностные, составные. Прогнозирование временных рядов	2			2			[1, 2, 5] о, [1, 3, 5] д	Опрос, защита лабораторной работы
7 8	Нейронные сети. Обработка естественного текста (NLP)	4			4			[1-3] о, [2, 5] д	Опрос, защита лабораторной работы, творческое задание
	ВСЕГО	8			8				Зачет

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Нейронные сети. [Электронный ресурс]. – Режим доступа: <http://statsoft.ru/home/textbook/modules/stneunet.html>. – Дата доступа: 15.05.2018.
2. Хайкин, Саймон. Нейронные сети. Полный курс, 2-е изд. / Саймон Хайкин. – М. : ООО «И. Д. Вильямс», 2018. – 1104 с. : ил.
3. Риз, Ричард. Обработка естественного языка на Java / Р. Риз. – М. : ДМК Пресс, 2015. – 246 с.
4. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. – М. : ДМК Пресс, 2015. – 400 с. : ил.
5. Рассел, М. Data mining. Извлечение информации из Facebook, Twitter, LinkedIn, Instagram / М. Рассел. – СПб. : Питер, 2020. – 464 с.

Перечень дополнительной литературы

1. Бокс, Дж. Анализ временных рядов. Прогноз и управление – Time series analysis. Forecasting and control : в 2 вып. Вып. 1 / Дж. Бокс, Г. Дженкинс ; пер. с англ. А. Л. Левшина ; под ред. [и с предисл.] В. Ф. Писаренко. – Москва : Мир, 1974. – 406
2. Вапник, В. Н. Теория распознавания образов : статистические проблемы обучения. – Москва : Наука, Главная редакция физико-математической лит., 1974.
3. Филипенков, Н. В. О некоторых вопросах анализа пучков временных рядов / Н. В. Филипенков, М. А. Петрова // Машинное обучение и анализ данных – 2014. – Т.1. – № 9. – С. 1215–1231
4. Чистяков, С. П. Случайные леса: обзор. / С. П. Чистяков // Труды Карельского научного центра РАН – 2013. – № 1. – С. 117–136.
5. Дюк, Вячеслав. Обработка данных на ПК в примерах / В. Дюк. – СПб. и др. : Питер, 1997. – 231с
6. Магнус, Я. Р. Эконометрика : начальный курс / Я. Р. Магнус, П. К. Катывшев, А. А. Пересецкий ; Академия народного хозяйства при Правительстве Российской Федерации. – М. : Дело, 1997. – 248 с.

ПЕРЕЧЕНЬ РЕКОМЕНДУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ И МЕТОДИКА ФОРМИРОВАНИЯ ИТОГОВОЙ ОЦЕНКИ

Диагностика результатов учебной деятельности по дисциплине «Интеллектуальный анализ данных» проводится, как правило, во время аудиторных занятий. Для диагностики используются устная и устно-письменная формы:

- экспресс-опрос на аудиторных занятиях;
- защита отчетов по заданиям для лабораторных работ;

Контрольные мероприятия проводятся в соответствии с учебно-методической картой дисциплины.

Формой текущей аттестации по дисциплине «Интеллектуальный анализ данных» учебным планом предусмотрен зачет.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Весовые коэффициенты, определяющие вклад текущего контроля знаний и текущей аттестации в рейтинговую оценку:

- отчеты по лабораторным работам – 50 %;
- выполнение творческого задания – 50 %.

Примерная тематика лабораторных занятий

1. Получение, предварительная подготовка, визуализация данных, оценка качества моделей (2 ч).
2. Отбор и выделение признаков, применение метод главных компонент (4 ч).
3. Методы классификации данных (4 ч).
4. Методы кластеризации (4 ч).
5. Применение иерархических моделей (4 ч).
6. Построение моделей для прогнозирования временных рядов (6 ч).
7. Нейронные сети для обработки изображений (6 ч).
8. Методы обработки естественного текста (6 ч).

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется **метод проектного обучения**, который предполагает:

- способ организации учебной деятельности студентов, развивающий актуальные для учебной и профессиональной деятельности навыки планирования, самоорганизации, сотрудничества и предполагающий создание собственного продукта;

- приобретение навыков для решения исследовательских, творческих, социальных, предпринимательских и коммуникационных задач.

При организации образовательного процесса используются **методы и приемы развития критического мышления**, которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения и письма; понимания информации как отправного, а не конечного пункта критического мышления.

При организации образовательного процесса используется **метод группового обучения**, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы:

- поиск (подбор) и обзор литературы и электронных источников по индивидуально заданной проблеме курса;
- научно-исследовательские работы;
- подготовка к участию в конференциях и конкурсах.

**ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ С ДРУГИМИ
ДИСЦИПЛИНАМИ СПЕЦИАЛЬНОСТИ**

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Приложения компьютерного моделирования	Кафедра Веб-технологий и компьютерного моделирования	Изменений нет	Оставить содержание учебной программы без изменений (протокол № 9 от 20.05.2020 г.)

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры *Веб-технологий и компьютерного моделирования* (протокол № __ от __ 202_ г.)

Заведующий кафедрой

докт. физ.-мат. наук, доцент

В.М. Волков

УТВЕРЖДАЮ

Декан факультета

докт. физ.-мат. наук, доцент

С.М. Босяков