

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

О. И. Чуприс

Регистрационный № УД 6229 /уч.



ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности второй ступени высшего
образования (магистратуры) с углубленной подготовкой специалиста:**

- 1-31 81 06 Веб-программирование и интернет-технологии
- 1-31 81 07 Математическое и программное обеспечение мобильных устройств
- 1-31 81 08 Компьютерная математика и системный анализ

2018 г.

Учебная программа составлена на основе ОСВО 1-31 81 08-2013 и учебных планов: №G31-227/уч. и G31-228/уч. от 10.04.2017г., G31з-230/уч. и №G31з-231/уч., G31-250/уч. и №G31з-266/уч. от 26.05.2017г.

СОСТАВИТЕЛИ:

Романчик Валерий Станиславович, заведующий кафедрой веб-технологий и компьютерного моделирования механико-математического факультета Белорусского государственного университета, кандидат физико-математических наук, доцент.

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

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой Веб-технологий и компьютерного моделирования
(протокол № 8 от 13.06.2018 г.);

Научно-методическим Советом Белорусского государственного университета
(протокол № 6 от 16.06.2018).

 Гуревский А. Н.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Дисциплина «Интеллектуальный анализ данных» является дисциплиной компонента учреждения высшего образования из цикла дисциплин специальной подготовки и адресована магистрантам 2-го года (3-й семестр) очной и заочной форм обучения.

Целью дисциплины «Интеллектуальный анализ данных» является ознакомление студентов 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 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.

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.

В результате изучения дисциплины «Интеллектуальный анализ данных» студент магистратуры будут сформированы следующие компетенции:

академические

- АК-2. Применять методологические знания и умения, обеспечивающие постановку и решение научно-педагогической и учебно-методической, научно-производственной, организационно-управленческой и инновационной деятельности.

- АК-3. Использовать междисциплинарный подход при решении проблем.

- АК-4. Применять технические устройства и компьютеры, использовать базы данных, пакеты прикладных программ и средства компьютерной графики для решения профессиональных задач.

- АК-5. Постоянно повышать свою квалификацию.

социально-личностные

- СЛК-5. Формировать и аргументировать собственные суждения и профессиональную позицию.

профессиональные

- ПК-2. Разрабатывать и использовать современное учебно-методическое обеспечение.

- ПК-3. Осваивать и внедрять в учебный процесс инновационные образовательные технологии.

- ПК-7. Квалифицированно проводить научные исследования в области математики и информационных технологий.

- ПК-8. Пользоваться глобальными информационными ресурсами.

- ПК-11. Взаимодействовать со специалистами смежных профилей.

- ПК-12. Применять современные методологии, формализованные языки и нотации, программные средства для построения и описания моделей процессов, данных, объектов.

- ПК-17. Осваивать и реализовывать управленческие инновации в сфере высоких технологий.

Дисциплина тесно связана с дисциплиной «Машинное обучение».

В соответствии с учебными планами специальностей на изучение дисциплины отводится:

Форма обучения	Срок обучения, лет	Курс	Семестр	Экзамен семестр	Зачет семестр	Всего часов	В том числе ауд.	Из них	
								лекций	лабораторных занятий
дневная	2	2	1	3		116	32	16	16
заочная	2,5	2	1	3		116	10	6	4

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 персептрон. 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
(очная форма обучения)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Литература	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	Предмет Data Mining	2			2		[6] о	Опрос, защита лабораторной работы
2	Снижение размерности	2			2		[1, 2] о	Опрос, защита лабораторной работы
3	Обучение с «учителем»	2			2		[5] о	Собеседование, защита лабораторной работы
4	Обучение без «учителя»	2			2		[5] о	Опрос, защита лабораторной работы
5	Модели: иерархические, вероятностные, составные	2			2		[2, 5] о, [5] д	Собеседование, защита лабораторной работы
6	Прогнозирование временных рядов	2			2		[1, 2] о [1, 3] д	Опрос, защита лабораторной работы
7	Нейронные сети	2			2		[3] о, [5] д	Опрос, защита лабораторной работы
8	Обработка естественного текста (NLP)	2			2		[4] о, [2] д	Опрос, защита лабораторной работы
	ВСЕГО	16			16			Экзамен

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ
(заочная форма обучения)

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

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

ЛИТЕРАТУРА

Основная

1. Дюк, Вячеслав. Обработка данных на ПК в примерах / В. Дюк. – СПб. и др. : Питер, 1997. – 231 с
2. Магнус, Я. Р. Эконометрика : начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий ; Академия народного хозяйства при Правительстве Российской Федерации. – М. : Дело, 1997. – 248 с.
3. Нейронные сети. [Электронный ресурс]. – Режим доступа: <http://statsoft.ru/home/textbook/modules/stneunet.html>. – Дата доступа: 15.05.2018.
4. Риз, Ричард. Обработка естественного языка на Java / Р. Риз. – М. : ДМК Пресс, 2015. – 246 с.
5. 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.
6. Чубукова, И. А. Data Mining : учеб. пособие / И. А. Чубукова. – Москва : Интернет-Университет Информационных Технологий ; : Бином. Лаборатория знаний, 2006. – 383 с. : ил.

Дополнительная

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. Хайкин, Саймон. Нейронные сети. Полный курс, 2-е изд. / Саймон Хайкин. – М. : ООО «И. Д. Вильямс», 2018. – 1104 с. : ил.
5. Чистяков, С. П. Случайные леса: обзор. / С. П. Чистяков // Труды Карельского научного центра РАН – 2013. – № 1. – С. 117–136.

ПЕРЕЧЕНЬ ИСПОЛЬЗУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ РЕЗУЛЬТАТОВ УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ

Диагностика результатов учебной деятельности по дисциплине «Интеллектуальный анализ данных» проводится в форме:

- опроса;
- собеседования;
- защиты лабораторных работ.

Задания лабораторных работ составляются согласно содержанию учебной программы.

МЕТОДИКА ФОРМИРОВАНИЯ ИТОГОВОЙ ОЦЕНКИ

Полученные студентом магистратуры количественные результаты учитываются как составная часть итоговой оценки по дисциплине в рамках рейтинговой системы.

Итоговая оценка формируется на основе трех документов:

- 1) Правила проведения аттестации.
- 2) Положение о рейтинговой системы БГУ.
- 3) Критерии оценивания студентов.

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 __ 201_ г.)

Заведующий кафедрой

канд. физ.-мат. наук, доцент

В.С. Романчик

УТВЕРЖДАЮ

Декан факультета

канд. физ.-мат. наук, доцент

Д.Г. Медведев