**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**УТВЕРЖДАЮ**

Проректор по учебной работе

и образовательным инновациям

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ О.И. Чуприс

(подпись) (И.О.Фамилия)

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2019 г.

Регистрационный № УД-\_\_\_\_\_ /уч.

**Технологии интеллектуального анализа данных**

**Учебная программа учреждения высшего образования**

**по учебной дисциплине для специальности:**

1-26 80 04 Менеджмент

Профилизация: Управление медиапроектами

2019 г.

Учебная программа составлена на основе ОСВО 1-26 80 04-2019 и учебных планов № Е26-014/уч. от 11.04.2019 г. и учебного плана для заочной формы получения образования № Е26з-016/уч. от 11.04.2019 г.

**Составители:**

И.А. Карачун, заведующий кафедрой корпоративных финансов экономического факультета БГУ, к.э.н., доцент

К.С. Бернат, преподаватель кафедры корпоративных финансов экономического факультета БГУ

Р.А. Гребнев, преподаватель кафедры корпоративных финансов экономического факультета БГУ

**Рецензенты[[1]](#endnote-1):**

В.О. Сувалов, ведущий специалист Управления регулирования ликвидности Национального банка Республики Беларусь

Е.И. Васенкова, доцент кафедры аналитической экономики и эконометрики БГУ, к.ф.-м.н., доцент

**РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:**

Кафедрой корпоративных финансов

(протокол № 9 от 26.06.19);

Научно-методическим Советом БГУ

(протокол № 5 от 28.06.19)

Заведующий кафедрой И.А. Карачун

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

Учебная программа учебной дисциплины «Технологии интеллектуального анализа данных» разработана для обучающихся по специальности 1-26 80 04 «Менеджмент» профилизации «Управление медиапроектами» в соответствии с образовательным стандартом специальности.

**Цели и задачи учебной дисциплины**

**Цель** учебной дисциплины – формирование у студентов теоретических знаний о технологиях интеллектуального анализа данных, знакомство с принципами работы по выявлению закономерностей в сложных данных, визуализации их в виде диаграмм и интерактивных средств просмотра.

**Задачи учебной дисциплины**:

1. ознакомление магистрантов с основами науки о данных;
2. формирование практических навыков работы с данными, решения прикладных задач анализа данных по специальности;
3. формирование у магистрантов навыков поиска информации.
4. овладение методами статистического анализа для систематизации и обработки результатов наблюдений случайных явлений, для выявления существующих статистических закономерностей.

**Место учебной дисциплины** в системе подготовки специалиста с высшим образованием (магистра).

Учебная дисциплина относится к циклу специальных дисциплин (компонент учреждения образования)

**Связи** с другими учебными дисциплинами, включая учебные дисциплины компонента учреждения высшего образования, дисциплины специализации и др.

В рамках учебной дисциплины «Технологии интеллектуального анализа данныхАнализ данных с использованием языка программирования R» расширяются и углубляются знания и практические навыки, необходимые при изучении учебной дисциплины «Медиаметрия и маркетинговые исследования СМИ».

**Требования к компетенциям**

Освоение учебной дисциплины «Технологии интеллектуального анализа данныхАнализ данных с использованием языка программирования R» должно обеспечить формирование следующих универсальных, углубленных профессиональных и специализированныхкомпетенций:

***универсальные*** компетенции:

* УК-1. Быть способным применять методы научного познания (анализ, сопоставление, систематизация, абстрагирование, моделирование, проверка достоверности данных, принятие решений и др.) в самостоятельной исследовательской деятельности, генерировать и реализовывать инновационные идеи.
* УК-9. Обладать навыками использования современных информационных технологий для решения научно-исследовательских и инновационных задач.

***углубленные профессиональные*** компетенции:

* УПК-1. Быть способным использовать управленческий потенциал информационных ресурсов и технологий, а также применять программные офисные инструменты для эффективного ведения бизнеса.

***специализированные*** компетенции:

* СК-2. Быть способным осуществлять анализ данных для решения экономических, управленческих, научно-исследовательских задач.

В результате освоения учебной дисциплины студент должен:

**знать**: основные методы и алгоритмы интеллектуального анализа и моделирования данных;

**уметь:** решать практические задачи вероятностными методами; строить вероятностные математические модели реальных экономических процессов;

**владеть:** приемами математической статистики при обработке и анализе экспериментальных данных для решения актуальных экономических задач.

**Структура учебной дисциплины**

Дисциплина изучается в 1 семестре. Всего на изучение учебной дисциплины «Технологии интеллектуального анализа данных» для очной формы получения образования отведено 108 часов, в том числе 50 аудиторных часов, из них: лекции – 18 часов, практические занятия – 32 часа.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма текущей аттестации – зачет.

**СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА**

**Тема 1. Бизнес-анализ в рамках анализа данных**

Аналитическая культура. Методологии управления данными. DAMA-DMBOK. Компоненты DAMA-DMBOK. Преимущества и недостатки методологии DAMA DMBOK. CMMI-DMMM. Компоненты CMMI DMMM. Итерационные методологии анализа данных: SEMMA, KDD, CRISP-DM.

**Тема 2. Введение в машинное обучение: теоретические аспекты моделирования данных**

Ключевые понятия и концепции машинного обучения. Модель, наблюдение, датасет. Количественные и качественные признаки. Обучение с учителем и без учителя. Понятия специфичности и чувствительности.

**Тема 3. Машинное обучение: обучение с учителем**

Понятие регрессии. Линейная регрессия. Логистическая регрессия. Дилемма смещения-дисперсии. Понятие классификации. Деревья решений. Композиционное обучение. Сэмплирование. Бутстрап. Бэггинг. Алгоритм случайного леса. Многоклассовая классификация. Гиперпараметры и кросс-валидация. Визуальный анализ.  ROC-кривая. PR-кривая. Lift-метрика, Gain. Кривая валидации. Кривая обучения. Регуляризация алгоритма. Методы отбора признаков.

**Тема 4. Машинное обучение. Обучение без учителя**

 Понятие кластеризации. Иерархические и плоские алгоритмы. Четкие и нечеткие алгоритмы. Метод ближайших соседей. Алгоритм word2vec. Понижение размерности. Метод главных компонент.

**УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ**

Дневная форма получения образования

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер раздела, темы | Название раздела, темы | Количество аудиторных часов | | | | | Количество часов  УСР | Форма контроля знаний |
| Лекции | Практические  занятия | Семинарские  занятия | Лабораторные  занятия | Иное |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | Бизнес-анализ в рамках анализа данных |  |  |  |  |  |  |  |
| 1.1 | Введение в анализ данных | 2 | 2 |  |  |  |  | Дискуссия |
| 1.2 | Методологии управления данными | 2 | 8 |  |  |  |  | Дискуссия |
| 2 | Введение в машинное обучение: теоретические аспекты моделирования данных |  |  |  |  |  |  |  |
| 2.1 | Основы практической статистики в рамках анализа данных |  | 4 |  |  |  |  | Проект, дискуссия |
| 2.2 | Ключевые понятия и концепции машинного обучения | 2 |  |  |  |  |  | Дискуссия |
| 3 | Машинное обучение: обучение с учителем |  |  |  |  |  |  |  |
| 3.1 | Регрессия: понятие, классификация, метрики | 2 | 2 |  |  |  |  | Проект, дискуссия |
| 3.2 | Классификация: понятие, деревья решений | 2 | 2 |  |  |  |  | Дискуссия |
| 3.3 | Композиционное обучение: сэмплирование, бутстрап, бэггинг. Алгоритм случайного леса. | 2 | 2 |  |  |  |  | Дискуссия, открытое эвристическое задание |
| 3.4 | Визуальный анализ. Регуляризация алгоритма. Методы отбора признаков. | 2 | 4 |  |  |  |  | Дискуссия, открытое эвристическое задание |
| 4 | Машинное обучение: обучение без учителя |  |  |  |  |  |  |  |
| 4.1 | Понятие кластеризации. Метод ближайших соседей. Алгоритм word2vec. | 2 | 4 |  |  |  |  | Доклад, дискуссия |
| 4.2 | Понижение размерности. Метод главных компонент. | 2 | 4 |  |  |  |  | Дискуссия |
|  | Итого | 18 | 32 |  |  |  |  |  |

**ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ**

**Перечень основной литературы**

1. Брюс, П. Практическая статистика для специалистов Data Science. 50 важнейших понятий / П. Брюс, Э. Брюс. – БХВ-Петербург, 2018. – 304 с.
2. Grolemund, G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data / Garrett Grolemund, Hadley Wickham. – O'Reilly, 2017. – 520 р.
3. Гвидо, С. Введение в машинное обучение с помощью Python: Пер. с англ. / Сара Гвидо, Андреас Мюллер. – М.: Вильямс, 2017. – 480 с.
4. Грас, Дж. Data Science. Наука о данных с нуля: Пер. с англ. / Джоэл Грас. – СПб.: БХВ-Петербург, 2019. – 336 с.
5. Маккинни, У. Python и анализ данных: Пер. с англ. / Уэс Маккинни. – М.: ДМК Пресс, 2015. – 482 с.

**Перечень дополнительной литературы**

1. Брюс, П. Практическая статистика для специалистов Data Science: Пер. с англ. / Питер Брюс, Эндрю Брюс. – СПб.: БХВ-Петербург, 2018. – 304 с.
2. Жерон, О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем.: Пер. с англ. / Орельен Жерон. – СПб.: ООО «Диалектика», 2018. – 688 с.

**Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки**

Оценка практической работы формируется на основе следующих критериев: корректность полученных результатов и их интерпретация, умение воспроизвести выполнение заданий, понимание практической применимости результатов работы, полнота ответов на вопросы. Итоговая оценка за практические работы рассчитывается путем усреднения оценок за запланированные к выполнению практические работы.

Оценка за проект формируется на основе следующих критериев: обоснованность используемых методов анализа данных, корректность трактовки полученных результатов, организация работы группы.

Формой текущей аттестации по дисциплине «Технологии интеллектуального анализа данных» учебным планом предусмотрен зачет.

**Примерный перечень практических заданий**

1. Методологии управления данными
2. Теоретические аспекты моделирования данных
3. Ключевые понятия и концепции машинного обучения
4. Регрессия: понятие, классификация, метрики
5. Классификация: понятие, деревья решений
6. Композиционное обучение: сэмплирование, бутстрап, бэггинг.
7. Алгоритм случайного леса.
8. Визуальный анализ.
9. Регуляризация алгоритма.
10. Методы отбора признаков.
11. Понятие кластеризации.
12. Метод ближайших соседей.
13. Алгоритм word2vec.
14. Понижение размерности.
15. Метод главных компонент.

**Список вопросов к зачету**

1. Аналитическая культура.
2. Методологии управления данными.
3. Компоненты DAMA-DMBOK.
4. Преимущества и недостатки методологии DAMA DMBOK.
5. Компоненты CMMI DMMM.
6. Итерационные методологии: SEMMA, KDD, CRISP-DM.
7. Ключевые понятия и концепции машинного обучения.
8. Модель, наблюдение, датасет.
9. Количественные и качественные признаки.
10. Обучение с учителем и без учителя.
11. Понятия специфичности и чувствительности.
12. Понятие регрессии.
13. Линейная регрессия.
14. Логистическая регрессия.
15. Дилемма смещения-дисперсии.
16. Понятие классификации.
17. Деревья решений.
18. Композиционное обучение.
19. Сэмплирование.
20. Бутстрап.
21. Бэггинг.
22. Алгоритм случайного леса.
23. Многоклассовая классификация.
24. Гиперпараметры и кросс-валидация.
25. Визуальный анализ.
26. ROC-кривая.
27. PR-кривая.
28. Lift-метрика, Gain.
29. Кривая валидации.
30. Кривая обучения.
31. Методы отбора признаков.
32. Понятие кластеризации.
33. Иерархические и плоские алгоритмы.
34. Четкие и нечеткие алгоритмы.
35. Метод ближайших соседей.
36. Алгоритм word2vec.
37. Понижение размерности.
38. Метод главных компонент.

**Описание инновационных подходов и методов к преподаванию   
учебной дисциплины**

При организации образовательного процесса используется следующие инновационные подходы и методы.

***1. Практико-ориентированный подход,*** который предполагает***:***

* освоение содержания через решение практических задач;
* приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
* ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;
* использованию процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

***2. 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.

***3. Метод учебной дискуссии,*** которыйпредполагает участие студентов в целенаправленном обмене мнениями, идеями для предъявления и/или согласования существующих позиций по определенной проблеме.

Использование метода обеспечивает появление нового уровня понимания изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения.

***4. Методы и приемы развития критического мышления,*** которые представляют собой систему, формирующую навыки работы с информацией в процессе чтения и письма; понимании информации как отправного, а не конечного пункта критического мышления.

**ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО**

|  |  |  |  |
| --- | --- | --- | --- |
| 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) |
| 1. |  |  |  |
| 2. |  |  |  |

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на \_\_\_\_\_/\_\_\_\_\_ учебный год

|  |  |  |
| --- | --- | --- |
| №  п/п | Дополнения и изменения | Основание |
|  |  |  |

Учебная программа пересмотрена и одобрена на заседании кафедры

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (протокол № \_\_\_\_ от \_\_\_\_\_\_\_\_ 201\_ г.)

(название кафедры)

Заведующий кафедрой

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ученая степень, ученое звание) (подпись) (И.О.Фамилия)

УТВЕРЖДАЮ

Декан факультета

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ученая степень, ученое звание) (подпись) (И.О.Фамилия)

1. Рецензенты указываются, как правило, только в учебных программах УВО для дисциплин государственного компонента, по которым отсутствуют типовые учебные программы. В этих случаях должно быть не менее двух рецензий, одна из которых внешняя, вторая – внешняя или внутренняя. [↑](#endnote-ref-1)