**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**

**Анализ данных с использованием языка программирования R**

**Учебная программа учреждения высшего образования**

**по учебной дисциплине для специальности:**

1-25 80 03 Финансы, налогообложение и кредит

Профилизация: Финансовые технологии и банковское дело

2019 г.

Учебная программа составлена на основе ОСВО 1-25 80 03-2019, учебного плана № Е25-011/уч. от 11.04.2019 г. и учебного плана для заочной формы получения образования № Е25з-012/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:**

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**

Программа учебной дисциплины «Технологии интеллектуального анализа данных» разработана для обучающихся по специальности 1-25 80 03 «Финансы, налогообложение и кредит» профилизации «Финансовые технологии и банковское дело» 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 модулю «Количественные методы в финансах и банковском деле» компонента учреждения высшего 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 «Финансовые технологии в цифровой экономике».

**Требования к компетенциям**

Освоение учебной дисциплины «Технологии интеллектуального анализа данныхАнализ данных с использованием языка программирования R» должно обеспечить формирование специализированной 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 90 часов, в том числе 42 аудиторных часов, из них: лекции – 18 часов, практические занятия – 24 часа.

Форма текущей аттестации – зачет в 1 семестре.

Дисциплина изучается во 2 семестре заочной формы получения образования. Всего на изучение учебной дисциплины отведено 90 часов, в том числе 14 аудиторных часов, из них: лекции – 8 часов, практические занятия – 6 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма текущей аттестации – зачет во 2 семестре.

**СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА**

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

Тема 1.1. Введение в анализ данных

Тема 1.2. Методологии управления данными

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

Тема 2.1. Основы практической статистики в рамках анализа данных

Тема 2.2. Ключевые понятия и концепции машинного обучения

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

Тема 3.1. Регрессия: понятие, классификация, метрики

Тема 3.2. Классификация: понятие, деревья решений

Тема 3.3. Композиционное обучение: сэмплирование, бутстрап, бэггинг. Алгоритм случайного леса.

Тема 3.4. Визуальный анализ. Регуляризация алгоритма. Методы отбора признаков.

**Раздел 4. Машинное обучение: обучение без учителя**

Тема 4.1. Понятие кластеризации. Метод ближайших соседей. Алгоритм word2vec.

Тема 4.2. Понижение размерности. Метод главных компонент.

**УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ**

Дневная форма получения образования

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер раздела, темы | Название раздела, темы | Количество аудиторных часов | Количество часовУСР | Форма контроля знаний |
| Лекции | Практическиезанятия | Семинарскиезанятия | Лабораторныезанятия | 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 |  |  |  |  |  | Дискуссия, опрос  |
| 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, реферат  |
| 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, контрольная самостоятельная работа  |
| 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 | 2 |  |  |  |  | Дискуссия, открытое эвристическое задание |
| 4 | Машинное обучение: обучение без учителя |  |  |  |  |  |  |  |
| 4.1 | Понятие кластеризации. Метод ближайших соседей. Алгоритм word2vec.  | 2 | 6 |  |  |  |  | Доклад, дискуссия, опрос |
| 4.2 | Понижение размерности. Метод главных компонент. | 2 | 4 |  |  |  |  | Дискуссия, дискуссия, контрольная самостоятельная работа |
|  | Итого | 18 | 24 |  |  |  |  |  |

**УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ**

Заочная форма получения образования

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер раздела, темы | Название раздела, темы | Количество аудиторных часов | Количество часовУСР | Форма контроля знаний |
| Лекции | Практическиезанятия | Семинарскиезанятия | Лабораторныезанятия | Иное |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | Введение в машинное обучение: теоретические аспекты моделирования данных |  |  |  |  |  |  |  |
| 1.1 | Основы практической статистики в рамках анализа данных | 1 |  |  |  |  |  | Дискуссия, опрос  |
| 1.2 | Ключевые понятия и концепции машинного обучения | 1 |  |  |  |  |  | Дискуссия, опрос  |
| 2 | Машинное обучение: обучение с учителем |  |  |  |  |  |  |  |
| 2.1 | Регрессия: понятие, классификация, метрики | 2 | 2 |  |  |  |  | Проект, дискуссия |
| 2.2 | Классификация: понятие, деревья решений | 2 | 2 |  |  |  |  | Дискуссия, контрольная самостоятельная работа  |
| ~~4~~ | Машинное обучение: обучение без учителя |  |  |  |  |  |  |  |
| 4.2 | Понижение размерности. Метод главных компонент. | 2 | 2 |  |  |  |  | Дискуссия, реферат |
|  | Итого | 8 | 6 |  |  |  |  |  |

**ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ**

**Перечень основной литературы**

1. Брюс, П. Практическая статистика для специалистов Data Science. 50 важнейших понятий / П. Брюс, Э. Брюс. – СПб.: БХВ-Петербург, 2018. – 304 с.
2. Гвидо, С. Введение в машинное обучение с помощью Python: Пер. с англ. / С. Гвидо, А. Мюллер. – М.: Вильямс, 2017. – 480 с.
3. Грас, Дж. Data Science. Наука о данных с нуля: Пер. с англ. / Дж. Грас. – СПб.: БХВ-Петербург, 2019. – 336 с.
4. Маккинни, У. Python и анализ данных: Пер. с англ. / У. Маккинни. – М.: ДМК Пресс, 2015. – 482 с.
5. Макшанов, А.В. Технологии интеллектуального анализа данных : учебное пособие / А.В. Макшанов, А.Е. Журавлев. – СПб.: Лань, 2018. – 208 с.
6. Grolemund, G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data / G. Grolemund, H. Wickham. – O'Reilly, 2017. – 520 р.

**Перечень дополнительной литературы**

1. Мусаев, А.А. Интеллектуальный анализ данных: учебное пособие / А.А. Мусаев. – СПб.: СПбГТИ(ТУ), 2018. – 176 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.
3. Станкевич, Л.А. Интеллектуальные системы и технологии: учебник и практикум для бакалавриата и магистратуры: / Л.А. Станкевич. – М: Юрайт, 2017. – 394 с.
4. Шешолко, В.К. Интеллектуальный анализ данных: пособие / В.К. Шешолко. – Мн.: АУПП РБ, 2019. – 114 с.
5. Гринчук, А.В. Интеллектуальный анализ данных: для студентов магистратуры по специальности 1-26 81 01 Бизнес-администрирование [конспект лекций] / А.В. Гринчук. – Мн.: ИБМТ БГУ, 2015. – 67 с.
6. Фрэнкс, Б. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики / Б. Фрэнкс. – М: Манн, Иванов и Фербер, 2014. – 349 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.

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:

* ответы на практических занятиях – 25 %;
* контрольные самостоятельные работы, проекты – 50 %;
* подготовка доклада/реферата – 25 %.

**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.

**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.

**Методические рекомендации по организации самостоятельной работы обучающихся**

В овладении знаниями учебной дисциплины важным этапом является самостоятельная работа студентов. Рекомендуется бюджет времени для самостоятельной работы в среднем 1,5–2 часа на 2-х часовое аудиторное занятие.

Основными направлениями самостоятельной работы студента являются:

* первоначально подробное ознакомление с программой учебной дисциплины;
* ознакомление со списком рекомендуемой литературы по учебной дисциплине в целом и ее разделам, наличие ее в библиотеке и других доступных источниках, изучение необходимой литературы по теме, подбор дополнительной литературы;
* изучение и расширение лекционного материала преподавателя за счет специальной литературы, консультаций;
* подготовка к практическим занятиям по специально разработанным планам с изучением основной и дополнительной литературы;
* подготовка к выполнению диагностических форм контроля (контрольные самостоятельные задания на компьютере, доклады/рефераты);
* подготовка к зачету.

**Примерный перечень 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. Дилемма смещения-дисперсии.
14. Понятие классификации.
15. Деревья решений.
16. Композиционное обучение.
17. Сэмплирование.
18. Бутстрап.
19. Бэггинг.
20. Алгоритм случайного леса.
21. Многоклассовая классификация.
22. Гиперпараметры и кросс-валидация.
23. Визуальный анализ.
24. ROC-кривая.
25. PR-кривая.
26. Lift-метрика, Gain.
27. Кривая валидации.
28. Кривая обучения.
29. Методы отбора признаков.
30. Понятие кластеризации.
31. Иерархические и плоские алгоритмы.
32. Четкие и нечеткие алгоритмы.
33. Метод ближайших соседей.
34. Алгоритм word2vec.
35. Понижение размерности.
36. Метод главных компонент.

**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) |
|  |  |  |  |

**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

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

 (название кафедры)

Заведующий кафедрой

к.э.н., доцент \_\_\_\_\_\_\_\_\_\_\_\_\_\_ И.А. Карачун

УТВЕРЖДАЮ

Декан факультета

к.ф.-м.н., доцент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.А. Королева