УТВЕРЖДАЮ

Руководитель Департамента

анализа данных, принятия решений

и финансовых технологий

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_В.И. Соловьев

29.11.2019 г.

ПРИЛОЖЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

**ПРЕДИКТИВНАЯ АНАЛИТИКА БОЛЬШИХ ДАННЫХ**

НАПРАВЛЕНИЕ ПОДГОТОВКИ: **09.04.03 «Прикладная информатика»**

НАПРАВЛЕННОСТЬ ПРОГРАММЫ МАГИСТРАТУРЫ: **«Обработка больших данных и разработка интеллектуальных приложений»**

ФОРМА ОБУЧЕНИЯ: **Очная**

ГОД ПРИЕМА: **2020**

Год утверждения программы: 2019

*Одобрено Департаментом анализа данных, принятия решений и финансовых технологий*

*Протокол № 6 от «29» ноября 2019 г*

**Содержание Приложения**

|  |  |
| --- | --- |
| **Наименование разделов РПД** | **стр.** |
| Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине | 2 |
| Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся | 4 |
| Учебно-тематический план | 4 |
| Содержание семинаров, практических занятий | 5 |
| Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы | 6 |
| Перечень вопросов, заданий, тем для подготовки к текущему контролю | 7 |
| Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине | 8 |
| Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем. | 14 |

**2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине**

|  |  |  |  |
| --- | --- | --- | --- |
| **Код компетенции** | **Наименование**  **компетенции** | **Индикаторы достижения компетенции** | **Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции** |
| **ДНК-1** | Способность автоматизировать сбор, подготовку, преобразование, загрузку и хранение данных из различных источников, а также управлять развитием БД для использования данных в интеллектуальных информационных системах | 1.Решает задачи и подготавливает регламентирующие документы по объединению данных различной структуры в единые форматы представления с целью последующей обработки и анализа интеллектуальными информационными системами.  2. Проектирует БД на основе моделей бизнес-процессов и алгоритмов работы для накопления информации при решении специализированных задач в различных прикладных областях.  3.Планрует и осуществляет работы по адаптации существующих и внедрению новых БД в различные информационные системы на основе современных и перспективных технологий.  4.Владеет практическим инструментарием по созданию, адаптации, сопровождению и миграции БД. | **Знать:** регламентирующие документы по объединению данных различной структуры в единые форматы  **Уметь:** представлять данные различной структуры в единые форматы с целью последующей обработки и анализа интеллектуальными информационными системами.  **Знать:** основные модели бизнес-процессов и алгоритмы работы для накопления информации при решении задач предикативной аналитики  **Уметь:** проектировать БД при решении задач предикативной аналитики  **Знать:** современные и перспективные технологии для адаптации БД  **Уметь:** осуществлять работы по адаптации существующих и внедрению новых БД в различные информационные системы на основе предикативной аналитике  **Знать:** основные инструментарии по созданию, адаптации, сопровождению и миграции БД при предикативной аналитике больших данных  **Уметь:** использовать предикативную аналитику больших данных для создания, адаптации, сопровождению и миграции БД |
| **ДКН-4** | Способность разрабатывать интеллектуальные информационные системы встроенной аналитики больших данных, с применением моделей машинного обучения и теории сложных сетей | 1.Владеет навыками использования технологии больших данных для создания интеллектуальных информационных систем и соответствующих решений.  2.Использует технологии больших данных при проектировании и реализации аналитических информационных систем в области экономики.  3. Обладает практическими навыками по применению технологий больших данных при создании интеллектуальных информационных систем | **Знать:** технологии больших данных для создания интеллектуальных информационных систем  **Уметь:** использовать технологии больших данных для создания интеллектуальных информационных систем  **Знать:** технологии больших данных при проектировании и реализации аналитических информационных систем в области экономики.  **Уметь:** использовать технологии больших данных при проектировании и реализации аналитических информационных систем в области экономики  **Знать:** технологии больших данных при создании интеллектуальных информационных систем  **Уметь:** использовать технологии больших данных при создании интеллектуальных информационных систем |

**4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся**

Общая трудоемкость дисциплины составляет 4 зачетных единицы.

Вид промежуточной аттестации: экзамен

Вид текущего контроля – контрольная работа.

*Очная форма обучения, 2020 г.*

|  |  |  |
| --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Модуль 4**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **4/144** | **144** |
| ***Контактная работа-***  ***Аудиторные занятия*** | **32** | **32** |
| *Лекции* | 8 | 8 |
| *Семинары, практические занятия* | 24 | 24 |
| **Самостоятельная работа** | **112** | **112** |
| Вид текущего контроля | Контрольная работа | Контрольная работа |
| Вид промежуточной аттестации | Экзамен | Экзамен |

**5.2. Учебно-тематический план**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | **Наименование тем (разделов) дисциплины** | **Трудоёмкость в часах** | | | | | | **Формы текущего контроля успеваемости** |
| **Всего** | **Аудиторная работа** | | | | **Самостоятельная работа** |
| Общая, в т.ч.: | Лекции | Семинары, практические занятия | Занятия в интерактивных формах |
| 1. | Задачи предиктивного моделирования и расширенная бизнес-аналитика | 36 | 8 | 2 | 6 | 4 | 28 | Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям. |
| 2. | Интеллектуальный анализ данных и его применение в цифровом бизнесе | 36 | 8 | 2 | 6 | 4 | 28 |
| 3. | Когнитивные подходы, системы и сервисы в цифровом бизнесе | 36 | 8 | 2 | 6 | 4 | 28 |
| 4. | Моделирование и оценка результатов предиктивного анализа | 36 | 8 | 2 | 6 | 4 | 28 |
|  | В целом по дисциплине | 144 | 32 | 8 | 24 | 16 | 112 | Контрольная работа. |
|  | Итого в % |  |  |  |  | 50% |  |  |

**5.3. Содержание семинаров, практических занятий**

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)** | **Формы проведения занятий** |
| Задачи предиктивного моделирования и расширенная бизнес-аналитика | Многофакторный анализ оттока клиентов в телекоммуникационной компании с использованием языка программирования Python  Предиктивное моделирование телемаркетинговой кампании банка с использованием  *Рекомендуемые источники: п.8, [1]; п. 9, [17]* | Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений |
| Интеллектуальный анализ данных и его применение в цифровом бизнесе | Построение модели поддержки принятия решений в области инвестиционных проектов на краудфандинговой платформе с использованием технологий расширенной аналитики  *Рекомендуемые источники: п.8, [4]* | Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений |
| Когнитивные подходы, системы и сервисы в цифровом бизнесе | Разработка многофакторной модели для оценки стоимости недвижимости в заданном регионе на основе гибридных подходов с использованием языка программирования Python.  *Рекомендуемые источники: п.8, [3]; п. 9 [4], [5]* | Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений |
| Моделирование и оценка результатов предиктивного анализа | Построение предиктивной модели оценки надежности заемщика на основании бизнес-кейса крупного коммерческого банка.  Создание предиктивной модели рейтинга мобильных приложений на площадке крупного агрегатора  *Рекомендуемые источники: п.9, [17]* | Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений |

**6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы**

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов, отводимых на самостоятельное освоение** | **Формы внеаудиторной самостоятельной работы** |
| Задачи предиктивного моделирования и расширенная бизнес-аналитика | Эволюция и тренды в развитии бизнес-аналитики. Классические и современные подходы к обработке данных. Хранилища данных и большие данные. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию. |
| Интеллектуальный анализ данных и его применение в цифровом бизнесе | Нейросети и машинное обучение в задачах цифрового управления. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию. |
| Когнитивные подходы, системы и сервисы в цифровом бизнесе | Когнитивные карты. Интеллектуальный поиск закономерностей динамических систем. Информационная цепочка добавленной стоимости в системе искусственного интеллекта (ИИ). Облачные когнитивные приложения и сервисы. | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию. |
| Моделирование и оценка результатов предиктивного анализа | Проверка гипотез. Статистический анализ данных. Моделирование на основе многокритериальных оценок. Развертывание и внедрение предиктивной модели. Методы, применяемые при проектировании моделей. Определение факторных и определяющих переменных. Оценка взаимного влияния факторов. Понятие мощности прогноза. Загрузка данных. Моделирование. Деревья решений. Сценарный анализ. Визуализация результатов. Подготовка бизнес-кейса для предиктивного моделирования | Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию. |

**6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю**

***Пример контрольной работы***

Построить автоматический распознаватель спама, т. е. устройство, которое бы предотвращало засорение электронного почтового ящика спамом, путем его распознавания и удаления.

Задача состоит в том, чтобы построить хорошую предиктивную (предсказательную) модель, на основе имеющихся данных.

Заметим, что нельзя дать точное определение спама и поэтому в каждом случае классификатор нужно настраивать индивидуально.

Данные для примера были собраны администраторами Hewlett-Packard Support.

Описание таблицы исходных данных

наблюдений 4601, из которых спам составлял 39.4% (1813 сообщений было спамом)

Общее число переменных: 58, из которых 57 непрерывных предикторов, и одна – категориальная (индикатор того, является сообщение спамом или нет).

Предикторы показывают, как часто появляется данное слово или символ в электронном письме. Переменные 55-57 показывают характеристики длины последовательности непрерывных последовательностей заглавных букв.

Описание переменных

Эти предикторные переменные характеризуют спам. Очевидно, можно использовать другие параметры.

48 непрерывных вещественных [1, 100] переменных типа word\_freq\_WORD = процентному отношению слов WORD в письме, т. е. 100\*(число появлений слова WORD в данном письме к общему числу слов).

6 непрерывных вещественных [1, 100] переменных типа char\_freq\_CHAR= процентному отношению появления символов CHAR в письме, к общему количеству символов в письме.

1 непрерывная вещественная [1, …] переменная типа capital\_run\_length\_average= средней длине непрерывной последовательности заглавных букв.

1 непрерывная целая [1, …] переменная типа capital\_run\_length\_longest= наибольшей длине непрерывной последовательности заглавных букв

1 непрерывная целая [1, …] переменная типа capital\_run\_length\_total= суммарное число заглавных букв в письме.

**7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2 «Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

**Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний**

|  |  |  |
| --- | --- | --- |
| **Код компетенции** | **Наименование компетенции** | **Примеры заданий для оценки индикаторов достижения компетенции** |
| **ДНК-1** | Способность автоматизировать сбор, подготовку, преобразование, загрузку и хранение данных из различных источников, а также управлять развитием БД для использования данных в интеллектуальных информационных системах | **1.Решает задачи и подготавливает регламентирующие документы по объединению данных различной структуры в единые форматы представления с целью последующей обработки и анализа интеллектуальными информационными системами.**  **Задание 1.**  Спроектируйте и разработайте БД для хранения информации с последующим использованием в интеллектуальной информационной системе для задач предиктивной аналитики. Представьте различные архитектуры БД (реляционные, нерялиционные). Подготовьте регламентирующие документы.  **2. Проектирует БД на основе моделей бизнес-процессов и алгоритмов работы для накопления информации при решении специализированных задач в различных прикладных областях.**  **Задание 2.**  Спроектируйте БД на основе моделей бизнес-процессов и алгоритмов работы для накопления информации для задачи кредитного скоринга.  **3.Планрует и осуществляет работы по адаптации существующих и внедрению новых БД в различные информационные системы на основе современных и перспективных технологий.**  **Задание 3.**  Внедрите БД «Рекомендательные банковские сервисы» в разработанную информационную систему с использованием минимум двух технологий.  **4.Владеет практическим инструментарием по созданию, адаптации, сопровождению и миграции БД.**  **Задание 4.**  Постройте аналитическую модель поддержки принятия решений в области сбытовой стратегии автопредприятия с учётом требований информационной безопасности |
| **ДКН-4** | Способность разрабатывать интеллектуальные информационные системы встроенной аналитики больших данных, с применением моделей машинного обучения и теории сложных сетей | **1.Владеет навыками использования технологии больших данных для создания интеллектуальных информационных систем и соответствующих решений.**  **Задание 1.**  Используя технологию обработки больших данных на основе нейросети, разработайте модуль интеллектуальной информационной системы, позволяющий распознавать пользователя системы.  **2.Использует технологии больших данных при проектировании и реализации аналитических информационных систем в области экономики.**  **Задание 2.**  Осуществив парсинг сайта avito.ru, создайте датасет по продажам двухкомнатных квартир по городу Москве (минимум 12 признаков, 20 тыс. объектов) и на его основе найдите наиболее важные признаки, влияющие на цены квартир.  **3. Обладает практическими навыками по применению технологий больших данных при создании интеллектуальных информационных систем**  **Задание 3.**  Разработать многофакторную модель для оценки стоимости недвижимости в заданном регионе на основе гибридных подходов с использованием Python |

***Примеры типовых заданий***

1. Многофакторный анализ оттока клиентов в телекоммуникационной компании с использованием Python.
2. Предиктивное моделирование телемаркетинговой кампании банка с использованием Python.
3. Исследование методов и построение моделей прогнозирования ухода сотрудников из компании.
4. Разработка моделей прогнозирования сбытовой деятельности фармацевтической компании.
5. Многофакторный анализ эффективности производства художественных фильмов в киноиндустрии с использованием Python.
6. Построение модели поддержки принятия решений в области инвестиционных проектов на краудфандинговой платформе с использованием технологий расширенной аналитики.
7. Прогнозирование результатов деятельности металлургического производства на основе предиктивного моделирования с использованием Python.
8. Многофакторный анализ качества пищевого продукта на основе классификационных моделей с использованием Python.
9. Построение аналитической модели поддержки принятия решений в области сбытовой стратегии автопредприятия с учётом экологичности двигателя и других факторов.
10. Построение предиктивной модели оценки надежности заемщика на основании бизнес-кейса крупного коммерческого банка.
11. Создание предиктивной модели рейтинга мобильных приложений на площадке крупного агрегатора с использованием Python.

***Примерные вопросы для подготовки к экзамену***

1. Определение предиктивной бизнес-аналитики
2. Примеры информационно-аналитических систем (ИАС)
3. Отличия в обработке структурированных и неструктурированных данных
4. Сравнительный анализ Descriptive, Predictive analytics, Prescriptive analytics
5. Современные тенденции в развитии цифрового бизнеса
6. Основные стадии эволюционного развития информационной бизнес-аналитики.
7. Системы реального времени (RTS) и предприятия реального времени (RTE) в цифровом бизнесе
8. Роль социальных медиа для развития информационной бизнес-аналитики
9. Лидирующие поставщики облачных платформ предиктивной аналитики
10. Архитектура платформ расширенной бизнес-аналитики
11. Понятие метаданных
12. Многомерное представление данных
13. Хранилища данных для цифрового управления
14. Понятие и основные технологии Больших данных
15. Основные задачи Data Mining
16. Примеры использования Data Mining в бизнесе
17. Технологии и задачи Text Mining
18. Цели и задачи Web Mining
19. Понятие социальной сети
20. Основные направления Social Mining
21. Область применения и принципы когнитологии
22. Отличия систем искусственного от систем естественного интеллекта
23. Определение NBICS-технологии и ее связь с когнитивной наукой
24. Понятие «сильного искусственного интеллекта»
25. Познавательная модель кибернетического «черного ящика»
26. Познавательная модель когнитивного «белого ящика»
27. Возможности когнитивных моделей при проектировании интеллектуальных систем
28. Условия обеспечения эффективности когнитивного моделирования
29. Когнитивные методы анализа в интеллектуальных системах
30. Принципы использования когнитивных карт
31. Ключевые направления развития когнитивных технологий
32. Технологии, применяемые в когнитивной бизнес-аналитике
33. Определение предиктивного анализа и предиктивной аналитики
34. Перечислите основные этапы процесса исследования данных
35. Примеры применения предиктивной аналитики в бизнесе
36. Ведущие поставщики систем предиктивной аналитики
37. Нейросети и их применение в прогнозировании
38. Решение задачи классификации на основе моделей нейронных сетей
39. Решение задачи классификации с помощью деревьев решений
40. Принципы проектирования предиктивных моделей

***Пример экзаменационного билета***

Федеральное государственное образовательное бюджетное учреждение

высшего профессионального образования

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ**

**ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Департамент анализа данных, принятия решений и финансовых технологий

Дисциплина: Предиктивная аналитика больших данных

Факультет: Прикладной математики и информационных технологий.

Форма обучения: очная

Направление подготовки: Прикладная информатика

Профиль: Обработка больших данных и разработка интеллектуальных приложений

Учебный 20\_\_/20\_\_\_год \_\_\_\_ модуль

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №

1. Вычисление относительной важности признаков. (30 баллов)
2. Задача. Решение задачи предиктивной аналитики в области экономики и финансов (30 баллов).

Подготовил

Заместитель руководителя

Дата \_\_\_\_\_\_\_\_\_\_\_\_

**11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.**

* 1. Комплект лицензионного программного обеспечения:

Windows, Microsoft Office;

Антивирус ESET Endpoint Security.

11.2 Современные профессиональные базы данных и информационные справочные системы:

Информационно-правовая система «Консультант Плюс»;

Информационно-правовая система «Гарант»;

Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

Cистема комплексного раскрытия информации «СКРИН» -http://www.skrin.ru

11.3. Сертифицированные программные и аппаратные средства защиты информации ‒ не предусмотрено

11.4. Эконометрический пакет R и интерфейс RStudio или другие системы компьютерной математики (например, MAXIMA или Wolfram A).