

**Федеральное государственное образовательное бюджетное учреждение
высшего образования «ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ
ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ» (Финансовый
университет)**

**Департамент анализа данных, принятия
решений и финансовых технологий**

Моисеев Г.В.

ОСНОВЫ БОЛЬШИХ ДАННЫХ И ЛОГИСТИКИ

Рабочая программа дисциплины
для студентов, обучающихся по направлению подготовки
38.03.02 «Менеджмент», профиль «Логистика»

Москва 2019

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент анализа данных, принятия
решений и финансовых технологий**

УТВЕРЖДАЮ

Проректор по развитию
образовательных программ

_____ Е.А. Каменева

24.12. 2019 г.

Моисеев Г.В.

**ОСНОВЫ БОЛЬШИХ ДАННЫХ И
ЛОГИСТИКИ**

Рабочая программа дисциплины
для студентов, обучающихся по направлению подготовки
38.03.02 «Менеджмент», профиль «Логистика»

*Рекомендовано Ученым советом
факультета прикладной математики и информационных технологий
(протокол № 21 от 17.12.2019 г.)*

*Одобрено Советом учебно-научного Департамента анализа данных,
принятия решений и финансовых технологий
(протокол № 7 от 17.12.2019 г.)*

Москва 2019

УДК 339 (073)

ББК 65.40

Рецензент: Макрушин С.В., к.э.н., доцент Департамента анализа данных, принятия решений и финансовых технологий **М 74 Моисеев Г.В.**

Основы больших данных и логистики. Рабочая программа дисциплины для студентов, обучающихся по направлению подготовки 38.03.02 «Менеджмент», профиль «Логистика» (программа подготовки бакалавра)– М.: Финансовый университет при Правительстве Российской Федерации, Департамент анализа данных, принятия решений и финансовых технологий, 2019. – 26 с.

Дисциплина «**Основы больших данных и логистики**» относится к Профильному блоку дисциплин по выбору (для 2018 года приема), к Модулю дисциплин по выбору, углубляющих освоение профиля «Логистика», направление подготовки 38.03.02 «Менеджмент» (для 2019, 2020 года приема).

В рабочей программе дисциплины определены ее цель, место в структуре ОП, требования к результатам освоения дисциплины, содержание программы, тематика практических занятий, формы самостоятельной работы, оценочные средства для текущего контроля и промежуточной аттестации, учебно-методическое и информационное обеспечение.

Учебное издание

Моисеев Георгий Викторович

Основы больших данных и логистики

Рабочая программа дисциплины

Компьютерный набор и верстка Г.В. Моисеев

Формат 60×90/16. Гарнитура Times New Roman Усл.п.л.

2,62. Изд. № - 2019. Тираж экз.

Заказ № _____

Отпечатано в Финансовом университете

□ Моисеев Г.В., 2019

□ Финансовый университет, 2019

СОДЕРЖАНИЕ

| | |
|---|----|
| 1. Наименование дисциплины | 4 |
| 2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине..... | 4 |
| 3. Место дисциплины в структуре основной образовательной программы | 7 |
| 4. Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся..... | 8 |
| 5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий..... | 8 |
| 5.1. Содержание дисциплины | 8 |
| 5.2. Учебно-тематический план | 11 |
| 5.3. Содержание семинаров, практических занятий..... | 12 |
| 6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине | 14 |
| 6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы | 14 |
| 6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю | 15 |
| 7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине | 18 |
| 8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины..... | 25 |
| 9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины..... | 26 |
| 10. Методические указания для обучающихся по освоению дисциплины..... | 27 |
| 11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем | 28 |
| 12. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине | 28 |

1. Наименование дисциплины

«Основы больших данных и логистики».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Дисциплина «Основы больших данных и логистики» обеспечивает инструментарий формирования следующих компетенций: ПКП-2, ПКП-3, ПКП-4.

| Код компетенции | Наименование компетенции | Индикаторы достижения компетенции¹ | Результаты обучения (владения², умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции |
|------------------------|---------------------------------|--|--|
|------------------------|---------------------------------|--|--|

¹ Заполняется при реализации актуализированных ОС ВО ФУ и ФГОС ВО3++

² Владения формулируются только при реализации ОС ВО ФУ первого поколения и ФГОС ВО 3+

| | | | |
|-------|--|---|--|
| ПКП-2 | Способность решать логистические задачи на основе современных технологий управления. | <p>1. Применяет методы корректной постановки логистических задач.</p> <p>2. Использует современные технологии управления в качестве инструмента повышения эффективности логистической деятельности на предприятиях.</p> | <p><u>Знать</u> методы математической постановки задач.</p> <p><u>Уметь</u> применить математические методы постановки задачи.</p> <p><u>Знать</u> перечень современных технологий управления и способы их применения.</p> <p><u>Уметь</u> применять инструменты повышения эффективности на логистических подразделениях</p> |
|-------|--|---|--|

| | | | |
|--|--|---|---|
| | | <p>1. Демонстрирует навыки в получении и анализе информации, необходимой для решения логистических задач в условиях цифровой экономики.</p> | <p><u>Знать</u> способы поиска информации логистической тематики</p> <p><u>Уметь</u> анализировать полученную информацию логистической направленности</p> |
|--|--|---|---|

| | | | |
|-------|---|--|---|
| ПКП-3 | Способность выявлять тенденции развития логистических процессов и проводить оценку их эффективности на основе современных технических средств. | <p>1. Применяет методы оценки тенденций развития современного общества, экономики и менеджмента с целью установления степени их влияния на развитие логистики.</p> <p>2. Использует современные подходы при сравнении нескольких вариантов решений при оценке эффективности логистических компаний.</p> <p>3. Демонстрирует навыки в создании эффективной логистической системы.</p> | <p><u>Знать</u> тенденции развития современного общества, экономики и менеджмента</p> <p><u>Уметь</u> определять степень возможного влияния современных тенденций на развитие логистики</p> <p><u>Знать</u> методы оценки эффективности логистических компаний</p> <p><u>Уметь</u> использовать современные технологии при выборе вариантов решений</p> <p><u>Знать</u> способы и средства создания современных логистических систем</p> <p><u>Уметь</u> применять на практике навыки создания логистических систем</p> |
| ПКП-4 | Способность формулировать предложения по повышению эффективности управления цепями поставок и материальными потоками на основе специальных программных продуктов. | <p>1. Применяет специальные программные продукты для принятия решений о разработке эффективной системы управления цепями поставок и материальными потоками.</p> <p>2. Использует оптимизаторы и другие методы поиска оптимальных решений,</p> | <p><u>Знать</u> системы управления цепями поставок и материальными потоками</p> <p><u>Уметь</u> применять специальные программные продукты поддержки принятия решений в логистике</p> <p><u>Знать</u> методы оптимизации и области их применения при</p> |

| | | | |
|--|--|--|--|
| | | <p>встроенные в специальные программные продукты, для принятия решения о формировании направлений развития логистики в компаниях различной отраслевой направленности.</p> <p>3. Демонстрирует навыки в создании эффективной системы управления цепями поставок и материальными потоками.</p> | <p>принятии решений в логистике</p> <p>Уметь пользоваться специальными программными продуктами, имеющими функцию оптимизаторов логистических решений</p> <p>Знать системы управления цепями поставок и материальными потоками</p> <p>Уметь создавать эффективные системы управления в логистике</p> |
|--|--|--|--|

3. Место дисциплины в структуре основной образовательной программы

Дисциплина «Основы больших данных и логистики» относится к Профильному блоку дисциплин по выбору (для 2018 года приема), к Модулю дисциплин по выбору, углубляющих освоение профиля «Логистика», направление подготовки 38.03.02 «Менеджмент» (для 2019, 2020 года приема).

В процессе изучения дисциплины происходит знакомство с основными понятиями, методологиями, моделями, методами, методиками и технологиями обработки больших данных, знания о которых будут положены в основу формирования профессиональных компетенций выпускника направления «Менеджмент». При этом студенты приобретают опыт практического использования изучаемых технологий в практических задачах, связанных со сферой логистических услуг.

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

Общая трудоёмкость дисциплины составляет 5 зачётных единиц.

Вид промежуточной аттестации – экзамен.

Вид текущего контроля – домашнее творческое задание.

Очно-заочная форма обучения 2018, 2019, 2020 г.г.

| Вид учебной работы по дисциплине | Всего (в з/е и часах) | Семестр 6 (в часах) |
|--|------------------------------|-----------------------------|
| Общая трудоёмкость дисциплины | 5 з.е. 180 ч. | 180 |
| <i>Аудиторные занятия- Контактная работа</i> | 66 | 66 |
| <i>Лекции</i> | 32 | 32 |
| <i>Семинары, практические занятия</i> | 34 | 34 |
| Самостоятельная работа | 114 | 114 |
| Вид текущего контроля | Домашнее творческое задание | Домашнее творческое задание |
| Вид промежуточной аттестации | Экзамен | Экзамен |

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. От данных к ценности: Введение в науки о данных

Основные определения: интеллектуальный анализ данных, большие данные, машинное обучение.

Методы и задачи интеллектуального анализа данных, машинного обучения и обработки больших данных.

Области применения методов и технологий интеллектуального анализа данных, машинного обучения и обработки больших данных.

Примеры задач машинного обучения: поиск информации в интернете, распознавание изображений, лиц, эмоций, пола, возраста, ..., распознавание речи, языка, эмоциональной окраски текстов, прогнозирование продаж, прогнозирование оттока клиентов, кредитный скоринг, рекомендательные системы и др.

Основные характеристики больших данных и их влияние на сбор, хранение, обработку и анализ данных (4V).

Критерии аналитических задач, решение которых предпочтительно с использованием технологий Big Data.

Цикл обработки данных: поиск данных, сбор данных, очистка данных, трансформация данных, интеллектуальный анализ данных, интерпретация и практическое применение результатов.

Статистические основы обработки данных: краткий обзор выборочного метода, методов описательной статистики, построения интервальных оценок и проверки гипотез.

Обзор современных технологий машинного обучения: R, Python, Spark, Microsoft Azure ML.

Тема 2. Практическое использование моделей классификации и регрессии

Принципы разработки и оценки систем машинного обучения. Основные классы моделей машинного обучения: классификация с обучением, регрессионный анализ, кластерный анализ, поиск аномалий.

Задача прогнозирования продаж. Модель множественной линейной регрессии. Методы оценки качества моделей регрессии. Коэффициент детерминации, средняя абсолютная ошибка предсказания, средняя относительная ошибка предсказания.

Задача кредитного скоринга. Модель логистической регрессии и ее реализация в R/Python/Azure ML. Методы оценки качества моделей классификации: доля правильных ответов, точность, полнота, F1, AUC. Цены ошибок первого и второго рода.

Модели деревьев и лесов решений и их компьютерная реализация. Усиление (бустинг) деревьев решений.

Нейронные сети и их компьютерная реализация.

Машины опорных векторов и их компьютерная реализация.

Технологии улучшения моделей машинного обучения. Подбор оптимальных параметров моделей. Важность подготовки данных. Генерация синтетических признаков. Работа с пропущенными данными. Работа с несбалансированными выборками.

Тема 3. Практическое использование моделей кластерного анализа и поиска аномалий

Кластерный анализ и его компьютерная реализация. Задача сегментирования потребителей. Иерархические агломеративные алгоритмы. Метод K-средних.

Методы оценки качества моделей кластерного анализа. Расстояния между кластерами, расстояния между объектами внутри кластеров. Характеристики центров кластеров.

Задача снижения размерности факторного пространства. Метод главных компонент и его компьютерная реализация. Геометрическое и

экономическое содержание метода главных компонент. Использование методов снижения размерности для улучшения качества моделей машинного обучения. Использование методов снижения размерности для выявления латентных факторов.

Поиск аномалий. Задача поиска мошеннических транзакций.

Тема 4. Анализ текстов, изображений, эмоций и рекомендательные системы Принципы анализа текстовой информации.

Принципы анализа графической информации.

Принципы анализа эмоциональной окраски текстов.

Принципы создания рекомендательных систем.

Интеллектуальные сервисы и чат-боты.

Перспективы развития систем обработки больших данных и машинного обучения.

Финансовые технологии, основанные на обработке данных и машинном обучении: интеллектуальные кредитные сервисы, интеллектуальные страховые сервисы, интеллектуальные сервисы интернета вещей.

Современные практические исследования Финансового университета в области обработки данных и машинного обучения: индикаторы смены состояний финансовых рынков, анализ сложных сетей в управлении городами, энергетике, на транспорте, предиктивный ремонт оборудования, системы скоринга.

5.2. Учебно-тематический план

Набор 2018, 2019, 2020 г.

| | | | | |
|--|--|-----------------------------|-------------------|--|
| | | Трудоёмкость в часах | | |
| | | Всего | Аудиторная работа | |

| № п/п | Наименование тем (разделов) дисциплины | | Общая, в т.ч.: | Лекции | Семинары, практические занятия | Занятия в интерактивных формах | Самостоятельная работа | Формы текущего контроля успеваемости |
|-------|--|-----|----------------|--------|--------------------------------|--------------------------------|------------------------|--|
| 1 | От данных к ценности: Введение в науку о данных | 36 | 16 | 8 | 8 | 8 | 20 | Практические работы (построение компьютерной системы визуализации данных, анализ конкретной ситуации и обсуждение результатов) |
| 2 | Практическое использование моделей | 36 | 16 | 8 | 8 | 8 | 20 | |
| | классификации и регрессии | | | | | | | |
| 3 | Практическое использование моделей кластерного анализа и поиска аномалий | 72 | 18 | 8 | 10 | 10 | 54 | |
| 4 | Анализ текстов, изображений, эмоций и рекомендательные системы | 36 | 16 | 8 | 8 | 8 | 20 | |
| | В целом по дисциплине | 180 | 66 | 32 | 34 | 34 | 114 | Домашнее творческое задание |
| | Итого | | | | | 51% | | |

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 |
|--|---|--------------------------|
| | | |

| | | |
|--|--|---|
| <p>Тема 1.</p> <p>От данных к ценности: Введение в науки о данных</p> | <p>Примеры задач машинного обучения Основные характеристики больших данных Цикл обработки данных Статистические основы обработки данных Обзор современных технологий машинного обучения</p> <p><i>Рекомендуемые источники: раздел 8 [2, 3]</i></p> | <p>Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 человека) и коллективное обсуждение решений</p> |
| <p>Тема 2.</p> <p>Практическое использование моделей классификации и регрессии</p> | <p>Задача прогнозирования продаж Задача кредитного скоринга Модели деревьев и лесов решений Технологии улучшения моделей машинного обучения</p> <p><i>Рекомендуемые источники: раздел 8 [1, 2, 4]</i></p> | <p>Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 человека) и коллективное обсуждение решений</p> |
| <p>Тема 3.</p> <p>Практическое использование моделей кластерного анализа и поиска аномалий</p> | <p>Кластерный анализ и его компьютерная реализация Методы оценки качества моделей кластерного анализа Задача снижения размерности факторного пространства Поиск аномалий</p> <p><i>Рекомендуемые источники: раздел 8 [1, 2, 4]</i></p> | <p>Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 человека) и коллективное обсуждение решений</p> |
| <p>Тема 4.</p> <p>Анализ текстов, изображений, эмоций и рекомендательные системы</p> | <p>Принципы анализа текстовой информации. Принципы анализа графической информации. Принципы анализа эмоциональной окраски текстов. Принципы создания рекомендательных систем. Интеллектуальные сервисы и чат-боты. Перспективы развития систем обработки больших данных и машинного обучения</p> <p><i>Рекомендуемые источники: раздел 8 [1, 2, 4]</i></p> | <p>Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 человека) и коллективное обсуждение решений</p> |

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

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

| Наименование разделов, тем входящих в дисциплину | Указание разделов и тем, отводимых на самостоятельное освоение обучающимися | Формы внеаудиторной самостоятельной работы |
|---|---|---|
| Тема 1. От данных к ценности: Введение в науки о данных | Обзор современных технологий машинного обучения: R, Python, Spark, Microsoft Azure ML. | Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление практической работы. |
| Тема 2. Практическое использование моделей | Технологии улучшения моделей машинного обучения. Подбор оптимальных параметров моделей. Важность подготовки данных. Генерация синтетических признаков. Работа с | Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и |
| классификации и регрессии | пропущенными данными. Работа с несбалансированными выборками. | оформление практических работ. |
| Тема 3. Практическое использование моделей кластерного анализа и поиска аномалий | Поиск аномалий. Задача поиска мошеннических транзакций. | Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление практической работы. |

| | | |
|---|--|--|
| тема 4. Анализ текстов, изображений, эмоций и рекомендательные системы | Современные практические исследования Финансового университета в области обработки данных и машинного обучения: индикаторы смены состояний финансовых рынков, анализ сложных сетей в управлении городами, энергетике, на транспорте, предиктивный ремонт оборудования, системы скоринга. | Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение и оформление практических работ. |
|---|--|--|

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

Примерная тематика домашнего творческого задания

1. Постройте систему визуализация данных из социальных сетей.
2. Постройте систему прогнозирования продаж.
3. Постройте систему кредитного скоринга.
4. Оптимизируйте существующую систему кредитного скоринга.
5. Проведите сегментирование потребителей.
6. Постройте систему рекомендательного сервиса.
7. Поведите анализ эмоциональной окраски записей в социальных сетях.
8. Постройте распределенные файловые системы;
9. Проведите модель вычислений MapReduce;
10. Организуйте SQL over BigData. Hive;
11. Постройте Beyond MapReduce. Spark;
12. Проведите машинное обучение на больших данных;
13. Выполните потоковую обработку данных;
14. Организуйте Key-value хранилища в больших данных.
15. Выберите все опции из ..., при которых HDFS останется в рабочем состоянии, и сохранится доступ ко всем данным. Описание кластера:

"Кластер HDFS состоит из 1 NameNode, 1 Secondary NameNode и 6 DataNode. Фактор репликации равняется 3"

16. Какие Вы видите проблемы в имплементации метода (transformation) семплирования (sample)? (приведен пример кода на PySpark)
17. Может ли HDFS работать если упала NameNode?
18. Решить задачу Top100 (WordCount) с помощью MapReduce
19. В рамках какой сущности Kafka хранит упорядоченный во времени поток событий?
20. Мы работаем с базой данных Cassandra. В настройках указано: фактор репликации - уровень консистенции - quorum. Сколько нод должны успешно ("success") отработать, прежде чем ответить на запрос пользователя?

В рамках домашнего творческого задания Вам нужно написать эссе о гипотетическом сервисе для хранения и обработки Больших Данных в логистике. Оно может относиться как к внутренним проектам логистики, так и любым внешним (астрономия, градостроение, сельское хозяйство, ...).

Вам необходимо:

1. описать проект
2. оценить финансовые затраты на вычислительные ресурсы

Объем эссе: 1-2 страницы формата А4 (1800-3600 знаков) Структура эссе, должна выглядеть следующим образом:

1. [10%] Опишите Ваш сервис для хранения / обработки Больших Данных. Для каких целей необходимо использовать в этой сфере технологии обработки Больших Данных?

2. [20%] Опишите конфигурации DataNode и NameNode? Конфигурация: характеристики CPU, RAM, HDD, число

вычислительных узлов кластера (DataNodes). При описании конфигурации учитывайте, что вычислительный узел падает один раз за 3 календарных года. Этот параметр необходим, чтобы оценить расходы на дополнительные серверы, которые будет необходимо докупать в течение года.

3. [20%] Какой фактор репликации используется в Вашем кластере? Одинаковым ли будет фактор репликации для всех файлов, хранящихся в HDFS? Ответ поясните.

4. [20%] Рассчитайте ежедневный объем нагрузки на кластер: сколько раз в день будет происходить загрузка имеющихся данных в память вычислительных узлов (для обработки / анализа)? Какой объем времени необходим для прочтения объема данных, равных объемам ежедневной нагрузки?

5. [30%] Выберите два облачных провайдера и оцените предполагаемую сумму для месячной поддержки вычислительного кластера.

Примерные задачи по практикуму

1. Распознавание лица с помощью нейронных сетей.
2. Влияние GDPR на сбор и обработку больших данных.
3. Разработка лекарственных препаратов с помощью искусственного интеллекта.
4. Влияние технологии смарт-контрактов на межорганизационные процессы.
5. Метод наименьших квадратов применительно к задаче линейной регрессии.
6. Логистическая регрессия.

7. Наивный классификатор Байеса.
8. Алгоритм k-means.
9. Алгоритм Априори.
- 10.Использование готовых решений анализа данных (Weka, Orange, и т.д.).
- 11.Визуализация данных с помощью Tableau.
- 12.Алгоритм k-means, реализация в рамках парадигмы Map Reduce.
- 13.Регуляризация метода наименьших квадратов.
- 14.Нейронная сеть.
- 15.Алгоритм SVM

Критерии бальной оценки формы текущего контроля успеваемости

Критерии бальной оценки различных форм текущего контроля успеваемости содержится в соответствующих методических рекомендациях Департамента анализа данных, принятия решений и финансовых технологий.

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, знаний и умений

| Код компетенции | Наименование компетенции | Примеры заданий для оценки индикаторов достижения компетенции |
|-----------------|--|--|
| ПКП-2 | Способность решать логистические задачи на основе современных технологий управления. | <p>1.Применяет методы корректной постановки логистических задач.</p> <p>Задание</p> <p>Примеры задач больших данных. Основные виды данных.</p> <p>2. Использует современные технологии управления в качестве инструмента повышения эффективности логистической деятельности на предприятиях.</p> <p>Задание</p> <p>Ключевые компетенции аналитика. Отличия BI от Data Science.</p> <p>3.Демонстрирует навыки в получении и анализе информации, необходимой для решения логистических задач в условиях цифровой экономики.</p> <p>Задание</p> <p>Постановка задачи регрессионного анализа. Линейная регрессия. Метод наименьших квадратов. Привести примеры использования регрессионного анализа.</p> |
| ПКП-3 | Способность выявлять тенденции развития логистических процессов и проводить оценку их эффективности на основе современных технических средств. | <p>1.Применяет методы оценки тенденций развития современного общества, экономики и менеджмента с целью установления степени их влияния на развитие логистики.</p> <p>Задание</p> <p>Отличия задачи классификации от задачи регрессии. Определение модели и алгоритма. Процесс обучения. Проблема переобучения. Регуляризация. Cross validation. Привести примеры использования алгоритмов классификации.</p> |

| | |
|--|---|
| | <p>2. Использует современные подходы при сравнении нескольких вариантов решений при оценке эффективности логистических компаний.</p> <p>Задание</p> <p>Постановка задачи кластеризации. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.</p> <p>3. Демонстрирует навыки в создании эффективной логистической системы.</p> <p>Задание</p> <p>Отличия построения ассоциативного правила от решающего правила задачи классификации. Привести примеры использования ассоциативных правил</p> |
|--|---|

| | | |
|-------|---|--|
| ПКП-4 | Способность формулировать предложения по повышению эффективности управления цепями поставок и материальными потоками на основе специальных программных продуктов. | <p>1. Применяет специальные программные продукты для принятия решений о разработке эффективной системы управления цепями поставок и материальными потоками.</p> <p>Задание</p> <p>Парадигма Map Reduce. Описать принцип работы. Нарисовать диаграмму. Перечислить слабые и сильные стороны. Обозначить области применимости</p> <p>2. Использует оптимизаторы и другие методы поиска оптимальных решений, встроенные в специальные программные продукты, для принятия решения о формировании направлений развития логистики в компаниях различной отраслевой направленности.</p> <p>Задание</p> <p>Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры использования визуализации</p> <p>3. Демонстрирует навыки в создании эффективной системы управления цепями поставок и материальными потоками.</p> |
| | | <p>Задание</p> <p>Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы над проектом.</p> |

Примеры типовых заданий

1. Определение больших данных, ключевые характеристики.

Примеры задач больших данных. Основные виды данных.

2. Роль аналитика по данным (Data Scientist). Ключевые компетенции аналитика. Отличия BI от Data Science.

3. Корреляция и регрессионный анализ. Коэффициент корреляции. Графическое представление. Постановка задачи регрессионного анализа. Линейная регрессия. Метод наименьших квадратов. Привести примеры использования регрессионного анализа.

4. Классификация. Признаковое описание объекта и таблица объект свойства. Постановка задачи. Отличия задачи классификации от задачи регрессии. Определение модели и алгоритма. Процесс обучения. Проблема переобучения. Регуляризация. Cross validation. Привести примеры использования алгоритмов классификации. Дополнительный вопрос: привести модель в линейной регрессии.

5. Кластеризация. Метрики. Матрица парных расстояний. Постановка задачи кластеризации. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.

6. Ассоциативные правила. Определение. Достоверность и поддержка. Отличия построения ассоциативного правила от решающего правила задачи классификации. Привести примеры использования ассоциативных правил.

7. Парадигма Map Reduce. Описать принцип работы. Нарисовать диаграмму. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.

8. Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры использования визуализации.

9. «Жизненный цикл» проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных. Перечислить

используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы над проектом. Перечислить основные роли исполнителей проекта.

10. Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии.

Примерные вопросы для подготовки к экзамену

1. Интеллектуальный анализ данных, большие данные и машинное обучение: основные понятия и методы.
2. Области применения методов и технологий интеллектуального анализа данных, машинного обучения и обработки больших данных. Примеры задач.
3. Основные характеристики больших данных и их влияние на сбор, хранение, обработку и анализ данных (4V).
4. Критерии аналитических задач, решение которых предпочтительно с использованием технологий Big Data.
5. Цикл обработки данных.
6. Статистические основы обработки данных.
7. Современные технологии машинного обучения.
8. Принципы разработки и оценки систем машинного обучения.
9. Основные классы моделей машинного обучения.
10. Модель множественной линейной регрессии.
11. Модель логистической регрессии.
12. Методы оценки качества моделей классификации.
13. Модели деревьев и лесов решений и их компьютерная реализация.
Усиление (бустинг) деревьев решений.

14. Нейронные сети и их компьютерная реализация.
15. Машины опорных векторов и их компьютерная реализация.
16. Технологии улучшения моделей машинного обучения.
17. Кластерный анализ. Методы оценки качества моделей кластерного анализа.
18. Методы снижения размерности и их использование.
19. Метод главных компонент. Его геометрическое и экономическое содержание.
20. Поиск аномалий.
21. Принципы анализа текстовой информации.
22. Принципы анализа графической информации.
23. Принципы анализа эмоциональной окраски текстов.
24. Принципы создания рекомендательных систем.
25. Интеллектуальные сервисы и чат-боты.
26. Перспективы развития систем обработки больших данных и машинного обучения.
27. Финансовые технологии, основанные на обработке данных и машинном обучении.

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: «Основы больших данных и логистики»
38.03.02 «Менеджмент»

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № ___

1. Основные классы моделей машинного обучения. (15 баллов)
2. Интеллектуальные сервисы и чат-боты. (15 баллов)
3. Задача (30 баллов).

Заместитель руководителя

Подготовил

ФИО _____

Дата _____

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины**а) основная:**

1. Миркин, Б.Г. Введение в анализ данных: учебник и практикум для бакалавриата и магистратуры / Б.Г. Миркин. - Москва: Юрайт, 2019. - 174 с. – Текст : непосредственный. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Юрайт, 2020. — 174 с. — (Высшее образование).—ЭБС Юрайт. — URL: <https://urait.ru/bcode/450262> (дата обращения: 18.06.2020). — Текст : электронный
2. Калинина, В.Н. Анализ данных: Компьютерный практикум: учебное пособие / В.Н. Калинина, В.И. Соловьев. — Москва: КНОРУС, 2017. — 166 с. — Текст: непосредственный. — То же. — ЭБС BOOK.ru. — URL: <https://www.book.ru/book/929386> (дата обращения: 18.06.2020). — Текст: электронный.

б) дополнительная:

1. Просто о больших данных / Д.Гурвиц, А. Ньюджент, Ф.Халпер, М.Кауфман. — Москва: Эксмо, 2015. — (Библиотека Сбербанка. Т. 58). – ЭБ Сбербанка. – URL:

<https://sberbankvip.alpinadigital.ru/book/8007> (дата обращения: 18.06.2020). - Текст : электронный.

2. Форман, Дж. Много цифр. Анализ больших данных при помощи Excel / Дж. Форман; перевод с английского А. Соколовой. — Москва: Альпина Паблишер, 2016. — 461 с. — ЭБС ZNANIUM.com.— URL: <http://znanium.com/catalog/product/551044> (дата обращения: 18.06.2020) .— Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Информационно-образовательный портал Финансового университета при Правительстве Российской Федерации <http://portal.ufrf.ru/>.
2. Сайт кафедры департамента анализа данных, принятия решений и финансовых технологий. http://fa.ru/dep/data_analysis/
3. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
4. *Applied Machine Learning* / Microsoft. — <https://www.edx.org/course/applied-machine-learning-microsoft-dat2033x-1>
5. *Data Science Essentials* / Microsoft. — <https://www.edx.org/course/datascience-essentials-microsoft-dat203-1x-3>
6. *Principles of Machine Learning* / Microsoft. — <https://www.edx.org/course/principles-machine-learning-microsoftdat203-2x-3>
7. Профессиональный ресурс по машинному обучению. — <https://machinelearning.ru>
8. Специализация «Машинное обучение и анализ данных» / МФТИ и Яндекс. — <https://www.coursera.org/specializations/machine-learningdata-analysis>
9. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)

10. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
11. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
12. Электронно-библиотечная система Znanium <http://www.znanium.com>
13. «Деловая онлайн библиотека» издательства «Альпина Паблишер» <http://lib.alpinadigital.ru/en/library>
14. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
15. Электронно-библиотечная система издательства «ЮРАЙТ» urait.ru
16. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
17. Воронцов, К.В. Машинное обучение [электронный ресурс]: курс лекций / К.В. Воронцов. — Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение
[\(курс лекций, К.В.Воронцов\)](#)
18. <http://alexanderdyakonov.narod.ru/intro2datamining.pdf>

10.Методические указания для обучающихся по освоению дисциплины

Самостоятельная работа студентов проходит аудиторно и внеаудиторно. Организации самостоятельной работы служит календарнотематический план изучения дисциплины. В этом плане указана тематика лекций и практических занятий.

При подготовке к лекции целесообразно предварительно познакомиться с ее содержанием по рекомендованным источникам и выделить наиболее трудные вопросы. Во время лекций следует конспектировать содержание лекции. После занятий следует провести работу с конспектом: отредактировать записи, оформить конспект. При

оформлении целесообразно выделять специальным образом названия тем и формулировки вопросов, основные определения, формулировки теорем и примеры. Сделанные записи нужно сопоставить с учебниками и учебными пособиями и в случае расхождений проконсультироваться с преподавателем.

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:

1. Windows, Microsoft Office;
2. Антивирус ESET Endpoint Security.

11.2. Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Гарант»
2. Информационно-правовая система «Консультант Плюс»
3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>
4. Система комплексного раскрытия информации «СКРИН» - <http://www.skrin.ru>

11.3. Сертифицированные программные и аппаратные средства защиты информации – не предусмотрены

11.4. R и RStudio

11.5. Anaconda и Python

11.6. Azure ML

12. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Для освоения дисциплины необходимо любое вычислительное средство – компьютер, смартфон или планшет. Предпочтительно

использование компьютера. При этом возможно использование компьютеров компьютерных классах университета.

Все изучаемые технологии доступны на личных устройствах студентов в любой точке, где доступна сеть Интернет.