

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 03.12.2024 21:29:49
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Северо-Западный институт управления – филиал РАНХиГС

Кафедра бизнес-информатики
(наименование кафедры)

УТВЕРЖДЕНО

Директор СЗИУ РАНХиГС
А.Д. Хлутков

ПРОГРАММА МАГИСТРАТУРЫ
Аналитическое обеспечение информационной безопасности

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ФТД.02 Предсказательная Аналитика
(индекс, наименование дисциплины, в соответствии с учебным планом)

38.04.05 Бизнес-информатика
(код, наименование направления подготовки (специальности))

Очная
(форма обучения)

Год набора – 2024

Санкт-Петербург, 2024

Автор–составитель:

Доктор военных наук профессор, заведующий кафедрой бизнес-информатики Наумов Владимир Николаевич

Заведующий кафедрой бизнес-информатики

Доктор военных наук, профессор Наумов Владимир Николаевич

В новой редакции РПД ФТД.02 «Предсказательная аналитика» одобрена протоколом № 10 заседания кафедры бизнес-информатики от 26.06.2024 г

СОДЕРЖАНИЕ

1.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы	4
2.	Объем и место дисциплины в структуре образовательной программы	4
3.	Содержание и структура дисциплины	5
4.	Материалы текущего контроля успеваемости обучающихся	6
5.	Оценочные материалы промежуточной аттестации по дисциплине	18
6.	Методические материалы для освоения дисциплины	19
7.	Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет"	20
	7.1. Основная литература	20
	7.2. Дополнительная литература	21
	7.3. Нормативные правовые документы и иная правовая информация	21
	7.4. Интернет-ресурсы	21
	7.5. Иные источники	22
8.	Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы	22

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1. Дисциплина ФТД.02 «Предсказательная аналитика» обеспечивает овладение следующими компетенциями.

Таблица 1.1

Код компетенции	Наименование компетенции	Код компонента компетенции	Наименование компонента компетенции
ПКс-3	Способен управлять работами по сопровождению и проектами по созданию (модификации) ИС, автоматизирующих задачи организационного управления и бизнес-процессы	ПКс-3.1	Способен управлять данными при решении задач бизнес-аналитики

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

Таблица 1.2

ОТФ/ТФ (при наличии профстандарта)/ профессиональные действия	Код компонента компетенции	Результаты обучения
Обоснование решений	ПКс-3.1	на уровне знаний: Знать: – методы анализа данных и машинного обучения; – возможности программных средств статистической обработки и предсказательного анализа данных современные информационные системы обработки и анализа данных
		на уровне умения: Уметь: – применять программные средства анализа данных, поддержки принятия решений для решения задач прогнозирования; – применять информационные технологии в объеме, необходимом для целей бизнес-анализа
		на уровне навыков: Владеть: - решения задач анализа наборов данных и временных рядов

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

Общая трудоемкость дисциплины составляет 1 зачетную единицу, 36 академ. часов/27 астрон. часов.

Таблица 2

Вид работы	Трудоемкость (акад/астр.часы)
Общая трудоемкость	36/27
Контактная работа с преподавателем	24/18
Лекции	12/9
Практические занятия	12/9
Самостоятельная работа	12/9
Контроль	-
Формы текущего контроля	Т, О, ПЗ

Форма промежуточной аттестации	Зачет
--------------------------------	-------

Место дисциплины в структуре ОП ВО

Дисциплина изучается в 1-м семестре 1-го курса. Дисциплина ФТД.02 «Предсказательная аналитика» относится к факультативным дисциплинам учебного плана по направлению 38.04.05 Бизнес-информатика. Преподавание дисциплины опирается на дисциплины программы бакалавриата «Эконометрика», «Анализ данных», «Теория вероятностей», «Математическая статистика».

В свою очередь она создаёт необходимые предпосылки для освоения программ таких дисциплин, как Б1.О.07 «Аналитическая поддержка принятия решений», Б1.В.ДВ.01.01 «Методы анализа данных», Б1.В.ДВ.01.02 «Методы машинного обучения и нейронные сети», Б1.В.09 «Интеллектуальный анализ данных, текстов и изображений».

Дисциплина закладывает теоретический и методологический фундамент для овладения умениям и навыками в ходе Б2.О.01(У) «Проектно-аналитическая практика» и Б2.О.02 (Н) «Научно-исследовательская работа».

Знания, умения и навыки, полученные при изучении дисциплины, используются студентами при выполнении выпускных квалификационных работ.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Таблица 3

№ п/п	Наименование тем	Объем дисциплины, час.					Форма текущего контроля успеваемости**, промежуточной аттестации* **	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий			СР		
			Л/ДОТ	ПЗ/ДОТ	КСР	СРО		СП
Тема 1	Анализ и визуализация данных в R и python	10	4	4		2		Т, О
Тема 2	Моделирование временных рядов	14	4	4		6		Т, О, ПЗ
Тема 3	Машинное обучение в анализе временных рядов	12	4	4		4		Т, О, ПЗ
Промежуточная аттестация		-						Зачет
Всего (акад./астр. часы):		36/27	12/9	12/9		12/9		

Примечание:

2* - консультация, не входящая в общий объем дисциплины

Используемые сокращения:

Л – занятия лекционного типа (лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся) ;

ПЗ – практические занятия (виды занятия семинарского типа за исключением лабораторных работ) ;

КСР – индивидуальная работа обучающихся с педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях (в том числе индивидуальные консультации) ;

СР – самостоятельная работа, осуществляемая без участия педагогических работников организации и (или) лиц, привлекаемых организацией к реализации образовательных программ на иных условиях;

СП – самопроверка;

СРО – самостоятельная работа обучающегося

Опрос (О), тестирование (Т), практическое задание (ПЗ)

3.2. Содержание дисциплины

Тема 1. Анализ и визуализация данных в R и python

Понятие анализа данных. Понятие «машинное обучение». Соотношение понятий BI, DM, AI.

Этапы типичного анализа данных. Data Mining. Жизненный цикл процесса анализа данных. Классификация методов Data Mining. Модели Data Mining. Понятие данные и знания. Классификация задач Data Mining. Пространственные данные, временные ряды, панельные данные. Общая характеристика моделей временных рядов.

Сравнительный анализ Python, R. Средства статистической обработки данных. Общая характеристика языка R. Графические средства языка. Среда моделирования R. Статистическая платформа JASP. Платформа Logiном. Парадигма программирования Low code.

Библиотеки и модули языков аналитики. Библиотека statsmodels, sklearn python. Библиотеки forecast, tseries, zoo. Библиотеки аккуратного программирования.

Тема 2. Моделирование временных рядов.

Регрессионный анализ временных рядов. Понятие фильтра и фильтрации. Экспоненциальное сглаживание. Сезонные и циклические составляющие временного ряда. Спектральный анализ временных рядов. Визуализация временных рядов. Коррелограмма. Периодограмма. Методология ARIMA.

Тема 3. Машинное обучение в анализе временных рядов.

Генерация и выбор признаков для временных рядов. Кластеризация временных рядов по исходным данным. DTW-расстояние. Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Пример кластеризации с помощью DTW-расстояния. Дендограммы. Метод K-средних. Понятие центроида. Профили кластеров. Кластеризация по описательным признакам.

4. Материалы текущего контроля успеваемости обучающихся

4.1. Формы контроля

В ходе реализации дисциплины «Предсказательная аналитика» используются следующие методы текущего контроля успеваемости обучающихся:

Таблица 3.1

Тема (раздел)	Формы (методы) текущего контроля успеваемости
Тема 1. Основы предсказательной аналитики	Тестирование, опрос
Тема 2. Моделирование временных рядов	Тестирование, опрос, задание
Тема 3. Машинное обучение в анализе временных рядов	Тестирование, опрос, задание

4. 2. Типовые материалы текущего контроля успеваемости обучающихся.

Типовые вопросы для опроса по теме 1

1. Дайте характеристику этапа ETL (Extracting Transforming and Loading).
2. Какие задачи решаются Data Mining?
3. Каково предназначение и средства разведочный Прогнозирование временных рядов? Дайте характеристику диаграммы «ящик с усами»
4. Дайте общую характеристику типов данных в задачах предсказательной аналитики.
5. Приведите примеры использования статистических пакетов для разведочного анализа.

6. Назовите и выполните сравнительный анализ графических средств анализа. Дайте характеристику биржевых диаграмм.

Тест по теме 1

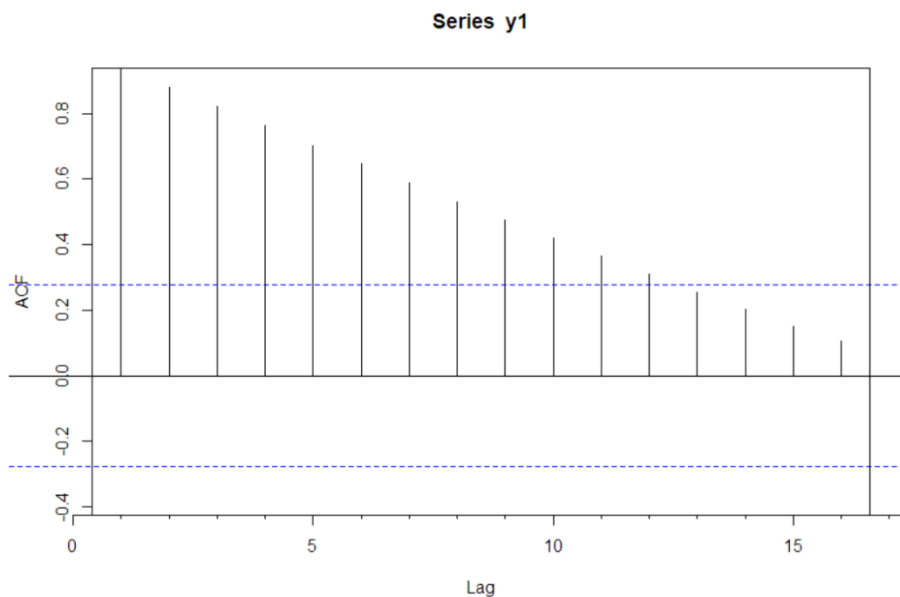
1. Информационные критерии. Что можно оценить с помощью информационного критерия Акайке?

1. стационарность
2. нестационарность
3. качество модели
4. ошибки аппроксимации модели
5. наличие аномалий

2. Компоненты модели временного ряда. Какие компоненты могут содержать модель временного ряда?

1. Тренд
2. случайная составляющая
3. сезонная составляющая
4. циклическая составляющая
5. аномальные компоненты
6. временная компонента
7. гармоника

3. Коррелограмма



Выберите модель временного ряда, коррелограмма для которого приведена выше

1. $t \leftarrow c(1:50)$
 $y1 \leftarrow -17.6 + 3.4 * t + rnorm(50, 0, 1.5)$

2. $t \leftarrow c(1:50)$
 $y1 \leftarrow -17.6 - 0.2 * \log(t) + rnorm(50, 0, 1.5)$

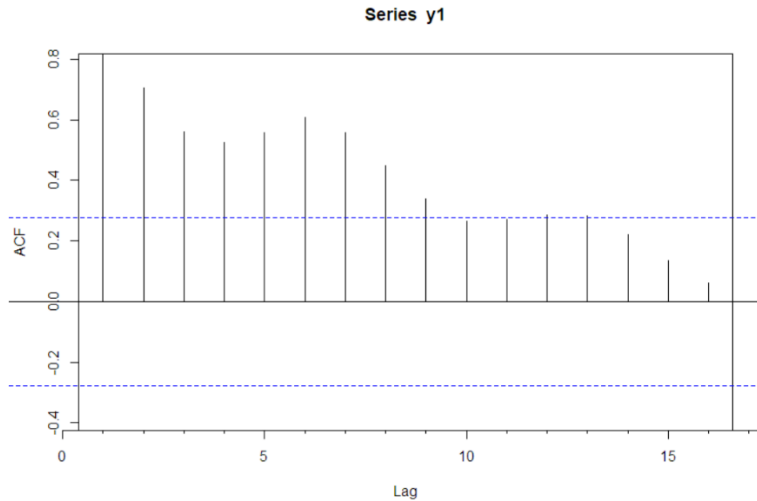
3. $t \leftarrow c(1:50)$
 $y1 \leftarrow -17.6 - runif(50, 0, 2)$

4. $t \leftarrow c(1:50)$
 $y1 \leftarrow -17.6 - rnorm(50, 0, 2) + runif(50, 0, 2)$

5. $t \leftarrow c(1:50)$
 $y1 \leftarrow -17.6 - \sin(1:50) + \text{runif}(50, 0, 2)$

6. $t \leftarrow c(1:50)$
 $y1 \leftarrow -17.6 + 3.4 * t + \text{rnorm}(50, 0, 1.5)$

4. Коррелограмма. Какими свойствами обладает модель временного ряда, если ее коррелограмма автокорреляционной функции имеет вид



1. наличие тренда
2. наличие сезонной составляющей
3. стационарностью
4. отсутствие тренда
5. эргодичностью

5. Отличия временных рядов от пространственных выборок. Выберите свойства, отличающие временные ряды от пространственной выборки

1. во временном ряду существенен порядок наблюдений, а в выборке не существен
2. уровни временного ряда измеряются в количественной шкале, а в пространственной выборке в любой шкале
3. члены временного ряда и наблюдения выборки независимы
4. члены временного ряда в общем случае статистически зависимы, а наблюдения независимы
5. члены временного ряда не являются одинаково распределенными, а наблюдения одинаково распределены
6. наблюдения выборки могут быть оценены с помощью статистических характеристик, а во временном ряду таких характеристик нет

6. Ошибка прогноза. Ошибка прогноза вычисляется по формуле

$$\sum |e_j|/N$$

где e_j - ошибка j -го наблюдения. Как называется данная ошибка?

1. MAE
2. MAPE
3. RSME
4. MASE
5. MPE

7. Ошибки прогноза. Ошибка прогноза вычисляется по формуле

$$\sum e_j^2/N$$

где e_j - ошибка j -го наблюдения. Как называется данная ошибка?

1. MAE
2. MSE
3. SSE
4. MAPE
5. MASE
6. RMSE

8. Планирование и прогнозирование. Чем отличается планирование от прогнозирования? Какими свойствами обладает планирование в отличие от прогнозирования?

1. носит директивный характер
2. имеет больший срок
3. имеет ресурсную обеспеченность
4. носит вероятностный характер
5. предусматривает несколько вариантов

9. Прогнозирование. Выберите свойства, которыми должно удовлетворять прогнозирование

1. вероятностный характер прогноза
2. носит характер исследования
3. горизонт прогноза не может быть больше периода ретроспекции
4. ошибки прогноза не могут быть больше 50%
5. обязательны несколько вариантов прогноза

10. модель временного ряда. Скрипт для синтеза уровней временного ряда имеет вид

```
y1<-17.6+3.4*t+rnorm(50,0,1.5)
```

```
y_ts1<-ts(y1,frequency = 12,start=c(2011,1))
```

Чему равен размер временного ряда?

10. Автокорреляционная функция

Укажите, какими свойствами обладает автокорреляционная функция

1. Для лага, равного нулю функция равна единице
2. Для лага, равного бесконечности функция равна нулю
3. Значения функции для лага, равного единице значимо отличаются от нуля
4. функция четная
5. функция нечетная

11. Критерии обнаружения выбросов

Укажите, какие критерии используются для выявления аномалий?

1. Критерий Граббса
2. Критерий Роснера
3. Критерий KPSS
4. Критерий Диксона
5. Критерий Бокса-Пирса

12. Критерий KPSS

Критерий KPSS предназначен для...

1. проверки нулевой гипотезы о том, что временной ряд не является TS-стационарным рядом
2. проверки нулевой гипотезы о том, что временной ряд является TS-стационарным рядом
3. проверки нулевой гипотезы о том, что временной ряд не является DS-

- стационарным рядом
4. проверки единичной гипотезы о том, что временной ряд не является DS-стационарным рядом
 5. проверки нулевой гипотезы о том, что временной ряд не является TS и DS-стационарным рядом

Типовые оценочные материалы по теме 2

Типовые вопросы для опроса по теме 2:

1. Что такое «сглаживание»? Его цели.
2. Что такое фильтр? Модель фильтра.
3. Метод скользящего среднего. Достоинства и недостатки метода скользящего среднего.
4. Что будет с результатами сглаживания, увеличить размер окна сглаживания?
5. Для чего используются взвешенные методы сглаживания?
6. В каких случаях используют методы адаптивного сглаживания?
7. Как оценить качество сглаживания?
8. Метод экспоненциального сглаживания. От чего зависит выбор параметра фильтра при экспоненциальном сглаживании?
9. Назовите особенности регрессионной модели прогнозирования.
10. Приведите примеры возможности использования регрессионных моделей в экономике.
11. Дайте общую характеристику метода наименьших квадратов.
12. Какие ограничения и допущения принимаются при использовании метода наименьших квадратов для оценки коэффициентов уравнения регрессии?
13. Как интерпретируются параметры простой линейной модели прогнозирования?

Тест по теме 2

№1. Тенденция временного ряда характеризует совокупность факторов, ...

1. оказывающих сезонное воздействие
2. не оказывающих влияние на уровень ряда
3. оказывающих единовременное влияние
4. оказывающих долговременное влияние и формирующих общую динамику изучаемого показателя

№2. Под автокорреляцией уровней временного ряда подразумевается _____ зависимость между последовательными уровнями ряда:

1. детерминированная
2. корреляционно–функциональная
3. корреляционная
4. функциональная

№3. Компонентами временного ряда являются:

1. коэффициент автокорреляции
2. лаг
3. тренд
4. циклическая (сезонная) компонента

№4. В стационарном временном ряде трендовая компонента ...

1. имеет линейную зависимость от времени
2. имеет нелинейную зависимость от времени
3. отсутствует
4. присутствует

№5. Период ряда. ежемесячные данные по рождаемости в г. Нью-Йорк, собранные в период

с января 1946 г. по декабрь 1959 г. представлены в виде временного ряда.

Пример заимствован из электронной книги A Little Book of R for Time Series и исходные данные можно загрузить с сайта проф. Роба Хиндмана (Rob J. Hyndman)

Скрипт для работы с ним имеет вид:

```
births <- scan("http://robjhyndman.com/tsdldata/data/nybirths.dat")
```

```
birthstimeseries <- ts(births, frequency=12, start=c(1946,1))
```

```
birthstimeseries
```

```
head(birthstimeseries)
```

Чему равен период сезонной составляющей временного ряда? Ответ дать числом

№6. Параметр фильтра

Укажите какой параметр используется в фильтре простого экспоненциального сглаживания

1. размер окна
2. альфа
3. бета
4. гамма
5. пси

№7. Простое экспоненциальное сглаживание

Какими свойствами обладает фильтр простого экспоненциального сглаживания?

1. Временной ряд считается стационарным. Тренд отсутствует
2. При определении сглаженного значения учитываются все предыдущие уровни временного ряда
3. Вес каждого предыдущего значения уровня временного ряда меньше на величину $\alpha(1-\alpha)$
4. Вес каждого предыдущего значения уровня временного ряда меньше на величину α
5. Прогнозируемое значение уровня временного ряда зависит от параметра фильтра

№8. Фильтр Брауна-Майера

Сколько уравнений задают поведение фильтра экспоненциального сглаживания Брауна-Майера, если сглаживаемый фильтр содержит кубический полиномиальный тренд?

1. 1
2. 2
3. 3
4. 4
5. 5

№9. Фильтр скользящего среднего

Какими свойствами обладает фильтр простого скользящего среднего?

сглаживание производится на центр окна сглаживания

каждое наблюдение имеет постоянный вес

все наблюдения зависят от его положения в окне фильтра

фильтр не позволяет вычислить значения для первых и последних значений временного ряда

сглаживание производится на правый край фильтра

№10. Аддитивная модель сглаживания

Ряд задан последовательностью $y \sim c(10,12,11,13,14,13,17,16,19,19,22, 24)$. Чему равно прогнозируемое значение уровня временного ряда на один шаг прогноза? При решении задачи прогнозирования использовать функцию ets. Использовать модель с аддитивным трендом и аддитивной моделью случайной составляющей. Ответ округлить с точностью одного знака после запятой

№11. Дисперсия фильтра

Дисперсия стационарного временного ряда равна 20. Чему равна дисперсия сглаженного ряда при использовании простого экспоненциального сглаживания, если его параметр равен 0,33? Ответ дать с точностью до целых

№12. Прогнозирование. Ряд задан последовательностью $y_t = c(10,12,11,13,14,13,17,16,19,19,22, 24)$. Чему равно прогнозируемое значение уровня временного ряда на один шаг прогноза? При решении задачи прогнозирования использовать функцию ets. Вид модели не указывать. Ответ округлить с точностью одного знака после запятой

№13. Фильтр взвешенного скользящего среднего WMA
Найти сглаженное значение фильтра WMA. если размер окна сглаживания равен 5, а значения ряда приведены в таблице

t	1	2	3	4	5
y	9	8	5	8	6

Сглаживание выполнить на правый край

Формула фильтра имеют вид
 $s = \text{Sum}(w_i * y_i) / \text{Sum}(w_i)$

№14. Фильтр сглаживания Хольта-Винтерса
Ряд задан последовательностью $y_t = c(10,12,11,13,14,13,17,16,19,19,22, 24)$. Решена задача фильтрации с помощью метода Хольта-Винтерса. Задачу решить в RStudio. При ее решении вид модели не задавать. Чему равно значение информационного критерия Акайке. Ответ дать с точностью одного знака после запятой

Практическое задание по теме 2

Исследовать временной ряд, проверить его стационарность. Решить задачу прогнозирования уровня временного ряда с использованием различных статистических методов. Набор данных взять на Сингапурском портале общедоступных данных (<https://data.gov.sg/search>), мэрии Парижа (<https://perma.cc/7V8Z-JZ4T>) – франкоязычный портал. При выполнении задания выполнить все этапы последовательного потокового анализа временных рядов с помощью статистических платформ и языков аналитики данных (Python). При решении задачи верификации модели использовать информационные критерии, показатели ошибок аппроксимации.

По количеству угроз нулевого дня построить и исследовать временной ряд, диаграмма которого приведена на рисунке (https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%A7%D0%B8%D1%81%D0%BB%D0%BE_%D0%BA%D0%B8%D0%B1%D0%B5%D1%80%D0%B0%D1%82%D0%B0%D0%BA_%D0%B2_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8_%D0%B8_%D0%B2_%D0%BC%D0%B8%D1%80%D0%B5#.2A_.D0.A7.D0.B8.D1.81.D0.BB.D0.BE_.D0.BA.D0.B8.D0.B1.D0.B5.D1.80.D0.B0.D1.82.D0.B0.D0.BA_.D0.BD.D0.B0_.D1.80.D0.BE.D1.81.D1.81.D0.B8.D0.B9.D1.81.D0.BA.D0.B8.D0.B5_.D0.BA.D0.BE.D0.BC.D0.BF.D0.B0.D0.BD.D0.B8.D0.B8_.D1.83.D0.B4.D0.B2.D0.BE.D0.B8.D0.BB.D0.BE.D1.81.D1.8C_.D0.B8_.D0.B4.D0.BE.D1.81.D1.82.D0.B8.D0.B3.D0.BB.D0.BE.C2.A0911.C2.A0.D1.82.D1.8B.D1.81).



Типовые оценочные материалы по теме 3

Типовые вопросы для опроса по теме 3

1. Что понимается под кластером? Назовите характеристики кластера. Что такое «центроид» кластера?
2. Дайте классификацию методов кластерного анализа. Приведите примеры их применения в практической жизни.
3. Зачем используются меры близости? Назовите методы определения близости между кластерами.
4. Когда применяется метод ближнего соседа, дальнего соседа? Сравните их.
5. Дайте характеристику метрик кластерного анализа.
6. Поясните содержание «дендрограммы» и организацию ее применения.
7. Что понимается под профилем кластера.
8. Использование статистических пакетов для решения задач кластерного анализа. Дайте характеристику метода k-средних

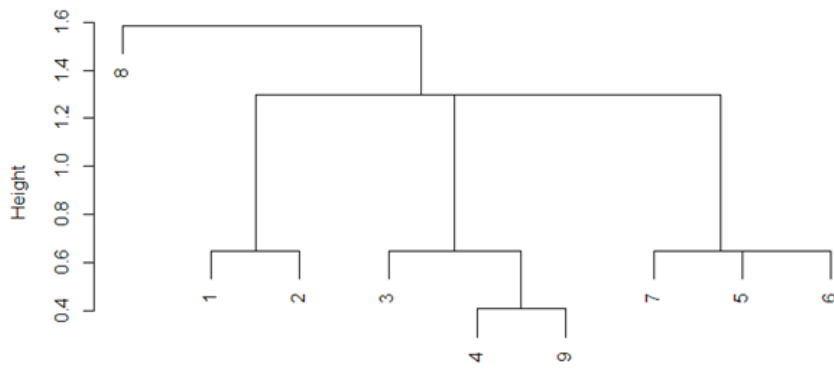
Тест по теме 3

1. Дендрограммы. Поставить соответствие между дендрограммами и используемыми методами для наблюдений, признаки которых заданы последовательностями
 $x \leftarrow c(2,3,4,5,6,7,8,3,5)$
 $y \leftarrow c(3,2,5,6,3,4,5,10,7)$.

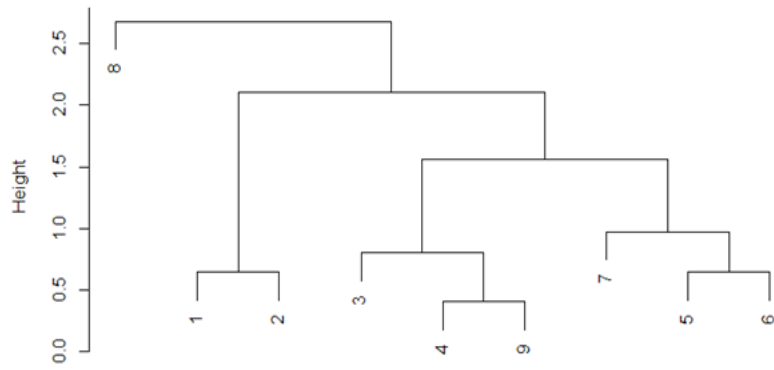
Cluster Dendrogram



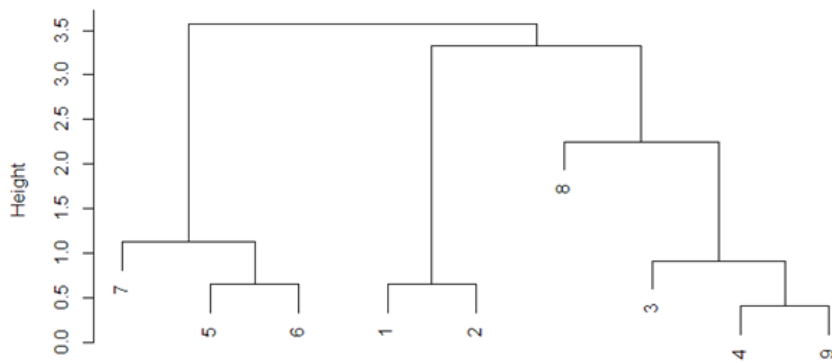
Cluster Dendrogram



Cluster Dendrogram



Cluster Dendrogram

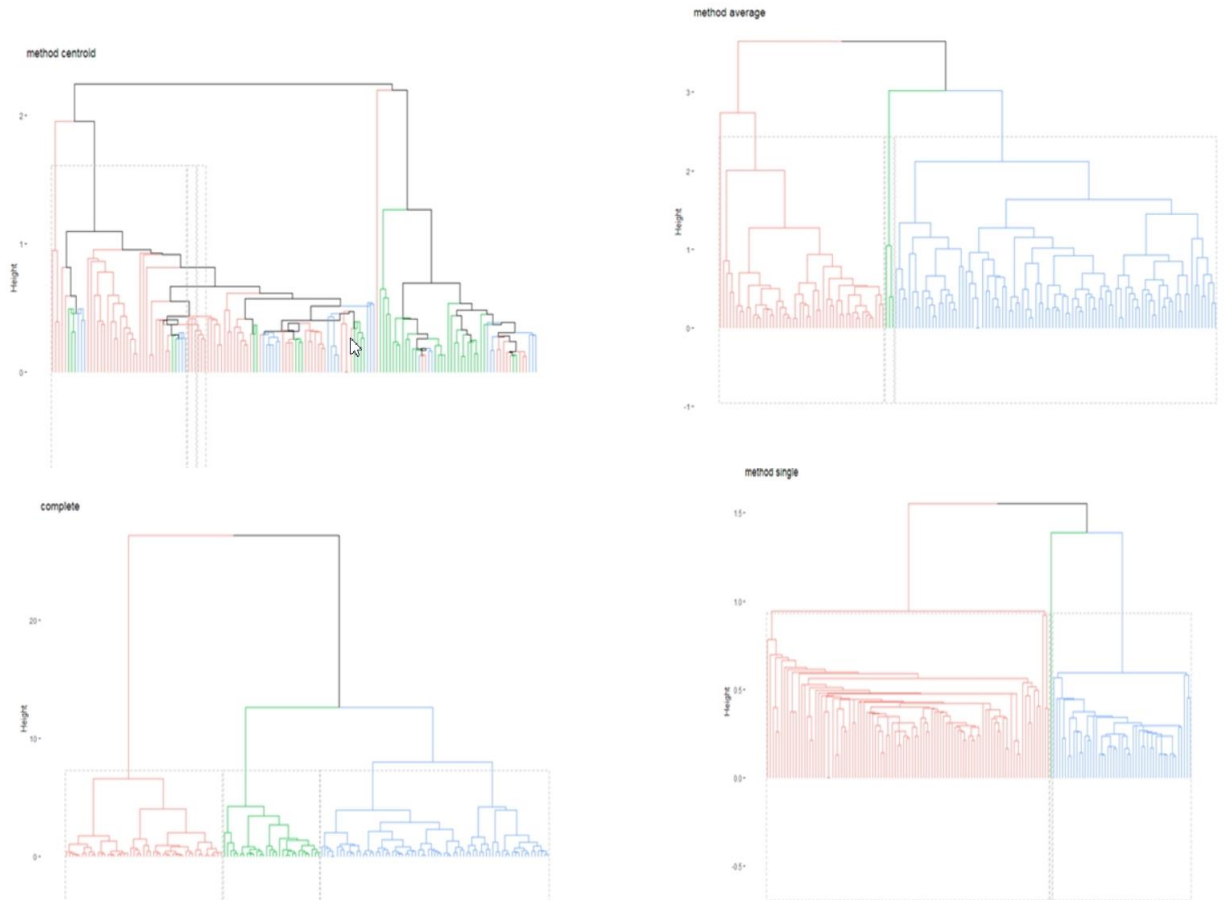


Варианты выбора:

1. центроидный
2. ближнего соседа
3. средней связи
4. полной связи
5. Варда

2. Иерархический кластерный анализ

При решении задачи кластерного анализа для набора данных iris различными методами построены дендрограммы с выделением кластеров прямоугольниками для заданного числа кластеров равного трем. Выберите лучший из методов, если дендрограммы имеют следующий вид



1 ближнего соседа

1. дальнего соседа
2. центроидный
3. средней связи
4. взвешенной средней связи

3. Методы иерархической кластеризации

Какой из методов кластерного анализа при анализе кандидатов на включения в кластер на текущем шаге использует результаты оценки дисперсий?

1. метод Варда
2. метод ближайшего соседа
3. метод полной связи
4. центроидный метод
5. метод невзвешенного попарного среднего

4. Внутрикластерная дисперсия

Используя RStudio решить задачу кластерного анализа методом k-средних для набора данных iris. При решении задачи использовать начальное значение генератора случайных чисел 1234. Задать максимальное число итераций равное 10. Число кластеров задать равным трем. Перед решением задачи выполнить стандартизацию значений параметров с помощью функции scale. Кластеризацию выполнять по четырем переменным набора данных (длине и ширине чашелистика и лепестка). Чему равно суммарное значение внутрикластерной дисперсии для всех трех кластеров?

5. Дендрограммы

Набор данных задан двумя последовательностями значений признаков.

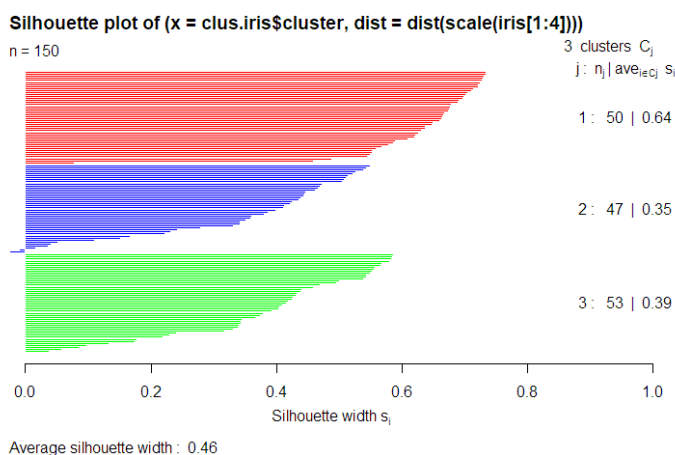
$x <- c(2, 3, 4, 5, 6, 7, 8, 3, 5)$

$y <- c(3, 2, 5, 6, 3, 4, 5, 10, 7)$.

Чему равно значение расстояния между четвертым и девятым наблюдениями, если все данными были стандартизованы с помощью функции scale, а также при построении дендрограммы использовать метод полной связи (дальнего соседа)? Ответ дать с точность до двух знаков после запятой

6. Диаграмма силуэтов

После решения задачи кластерного анализа для набора данных iris получена диаграмма силуэтов, которая имеет вид:



Укажите номер кластера, качество формирования которого наилучшее при ее оценке методом силуэтов? Номер лучшего кластера указать числом

7. метод k-средних

Используя RStudio решить задачу кластерного анализа методом k-средних для набора данных iris. При решении задачи использовать начальное значение генератора случайных чисел 1234. Задать максимальное число итераций равное 10. Число кластеров задать равным трем. Перед решением задачи выполнить стандартизацию значений параметров с помощью функции scale. Кластеризацию выполнять по четырем переменным набора данных (длине и ширине чашелистика и лепестка). Чему равно число наблюдений, попавших во второй кластер?

8. Метрики расстояния

Чему равно Евклидово расстояние между объектами, характеризуемыми двумя признаками, если известны значения признаков

2	4
3	5

Ответ дать с точностью до двух знаков после запятой

9. Профили кластеров

Задача кластерного анализа для набора данных iris решена в SPSS. Задача решалась методом двухэтапного кластерного анализа. В результате ее решения для заданного числа кластеров равного трем получена таблица профилей кластеров, имеющая вид:

Кластер		Центриды							
		SEPALL		SEPALW		PETALL		PETALW	
		Среднее	Станд. отклонения	Среднее	Станд. отклонения	Среднее	Станд. отклонения	Среднее	Станд. отклонения
1		5,0060	,35249	3,4280	,37906	1,4620	,17366	,2460	,10539
2		6,1742	,59177	2,8333	,29940	4,8065	,76298	1,6366	,40908
3		7,4286	,41519	3,3857	,34847	6,2286	,35456	2,2000	,25820
Объединенный		5,8433	,82807	3,0573	,43587	3,7580	1,76530	1,1993	,76224

Проанализировав данную таблицу, укажите номер кластера, качество которого наихудшее. В ответе номер указать числом

10. Расстояние Манхеттена

Чему равно расстояние Манхеттена между объектами, характеризуемыми двумя признаками, если известны значения признаков

2 4

3 5

Ответ дать с точностью до целых

Практическое задание по теме 3.

С помощью интерактивных карт киберугроз построить и исследовать временные ряды для различных вирусов.

Проанализировать три наиболее зараженные страны на текущий момент времени. <https://cybermap.kaspersky.com/ru/stats#country=213&type=OAS&period=m>.

Кластеризовать страны по количеству атак обнаруженных источниками данных: OAS, IDS, VUL, ODS, WAV, KAS, MAV

OAS (On-Access Scan) - автоматическая проверка. Показывает поток данных по вредоносным программам, обнаруженным во время открытия, копирования, запуска или сохранения файлов.

IDS (Intrusion Detection Scan). Система обнаружения вторжений показывает поток данных по обнаруженным сетевым атакам.

VUL (Vulnerability Scan). Поиск уязвимостей показывает поток данных по обнаруженным уязвимостям.

ODS (On Demand Scanner). Проверка по требованию показывает поток данных по вредоносным программам, возникающий, когда пользователь вручную выбирает "Просканировать компьютер" в меню.

WAV (Web Anti-Virus). Веб-антивирус показывает поток данных по вредоносным программам, обнаруженным при открытии HTML-страниц веб-сайтов, а также при загрузке файлов. Веб-антивирус проверяет порты, указанные в его настройках.

KAS (Kaspersky Anti-Spam) Касперский Анти-Спам показывает подозрительный и нежелательный почтовый трафик, обнаруженный с помощью технологий репутационной фильтрации «Лаборатории Касперского»

MAV (Mail Anti-Virus) . Почтовый антивирус показывает поток данных по вредоносным программам, обнаруженным среди новых объектов в почтовых приложениях. Почтовый антивирус проверяет входящие сообщения и запускает автоматическую проверку при сохранении

5. 5.Оценочные материалы промежуточной аттестации по дисциплине

5.1 Формы промежуточного контроля

Зачет проводится по итогам выполнения заданий текущего контроля.

5.2 Оценочные материалы промежуточной аттестации

Компонент компетенции	Промежуточный/ключевой индикатор	Критерий оценивания
ПКс-3.1	Управляет данными при решении задач бизнес-аналитики	Использует информацию, методы и программные средства ее сбора, обработки и анализа, в том числе с использованием интеллектуальных методов. Самостоятельно решает задачи предиктивной аналитики при анализе временных рядов

Для оценки сформированности компетенций, знаний и умений, соответствующих данным компетенциям, используются контрольные вопросы, а также задачи.

Типовые оценочные материалы промежуточной аттестации

Вопросы к зачету по дисциплине «Предсказательная аналитика»

1. Системы поддержки принятия решений. Хранилища данных.
2. Понятие «Анализ данных». Методы и средства анализа данных. Общая классификация.
3. Понятие «машинное обучение». Классификация методов машинного обучения.
4. Проблема больших данных. Средства работы с большими данными.
5. Размерностные модели. OLAP-куб. Таблица размерностей. Таблица фактов.
6. Сравнительный анализ OLAP и OLTP-систем.
7. Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
8. Этапы анализа данных. KDD.
9. Data Mining. Средства обработки Data Mining
10. Элементы математической статистики. Описательная статистика. Операции агрегирования данных.
11. Графические средства анализа. Диаграмма рассеяния. Гистограмма.
12. Начальные этапы KDD. ETL. Средства очистки и трансформации данных.
13. Классификация метод предобработки и очистки данных.
14. Методы борьбы с аномалиями. Ящичная диаграмма. Статистические критерии выявления аномалий
15. Общая характеристика задач кластерного анализа.
16. Метрики кластерного анализа.
17. Методы определения близости между кластерами.
18. Иерархическая кластеризация. Дендрограмма.
19. Метод к-средних.
20. Кластеризация временных рядов
21. Общая характеристика языка R.
22. Объекты языка R.
23. Работа с данными языка R.
24. Решение задач разведывательного анализа в R.

25. Построение статистических моделей в R.
26. Понятие временного ряда.
27. Модели временных рядов.
28. Решение задач сглаживания временных рядов.
29. Методология ARIMA.
30. Методика построения моделей в ARIMA.
31. Примеры построения моделей в методологии ARIMA

Шкала оценивания

Оценка результатов производится на основе Положения о текущем контроле успеваемости обучающихся и промежуточной аттестации обучающихся по образовательным программам среднего профессионального и высшего образования в федеральном государственном бюджетном образовательном учреждении высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», утвержденного Приказом Ректора РАНХиГС при Президенте РФ от 30.01.2018 г. № 02-66 (п.10 раздела 3 (первый абзац) и п.11), а также Решения Ученого совета Северо-западного института управления РАНХиГС при Президенте РФ от 19.06.2018, протокол № 11.

Зачет

На «зачтено» оцениваются ответ, в котором системно, логично и последовательно изложен материал на все поставленные вопросы. Кроме того, студент должен показать способность делать самостоятельные выводы, комментировать излагаемый материал. При этом допускаются некоторые затруднения с ответами, например, затруднения с примерами из практики, затруднения с ответами на дополнительные вопросы.

«Не зачтено» ставится в случае, когда студент не знает значительной части учебного материала, допускает существенные ошибки; знания носят бессистемный характер; на большинство дополнительных вопросов даны ошибочные ответы; ответ дается не по вопросу.

6. Методические материалы для освоения дисциплины

Рабочей программой дисциплины предусмотрены следующие виды аудиторных занятий: лекции, практические занятия. На лекциях рассматриваются наиболее сложный материал дисциплины. Для развития у магистрантов креативного мышления и логики в каждой теме учебной дисциплины предусмотрены теоретические положения, инструментальные средства, а также примеры их использования при решении задач предиктивной аналитики. Кроме того, часть теоретического материала предоставляется на самостоятельное изучение по рекомендованным источникам для формирования навыка самообучения.

Практические занятия предназначены для самостоятельной работы магистрантов по решению конкретных задач. Каждое практическое занятие сопровождается заданиями, выдаваемыми магистрантам для решения во внеаудиторное время.

Для работы с печатными и электронными ресурсами СЗИУ имеется возможность доступа к электронным ресурсам. Организация работы магистрантов с электронной библиотекой указана на сайте института (странице сайта – «Научная библиотека»).

7. Методические указания для обучающихся по освоению дисциплины

Обучение по дисциплине «Предсказательная аналитика» предполагает изучение курса на аудиторных занятиях (лекции, практические работы) и самостоятельной работы обучающихся. Семинарские занятия дисциплины «Предсказательная аналитика»

предполагают их проведение в различных формах с целью выявления полученных знаний, умений, навыков и компетенций с проведением контрольных мероприятий. С целью обеспечения успешного обучения обучающийся должен готовиться к лекции, поскольку она является важнейшей формой организации учебного процесса, поскольку:

- знакомит с новым учебным материалом;
- разъясняет учебные элементы, трудные для понимания;
- систематизирует учебный материал;
- ориентирует в учебном процессе.

Подготовка к лекции заключается в следующем:

- внимательно прочитайте материал предыдущей лекции;
- узнайте тему предстоящей лекции (по тематическому плану, по информации лектора);
- ознакомьтесь с учебным материалом по рекомендуемой литературе;
- постарайтесь уяснить место изучаемой темы в своей профессиональной подготовке;
- запишите возможные вопросы, которые вы зададите лектору на лекции.

Подготовка к практическим занятиям:

- внимательно прочитайте материал лекций, относящихся к данному семинарскому занятию, ознакомьтесь с учебным материалом;
- ответьте на контрольные вопросы по семинарским занятиям, готовьтесь дать развернутый ответ на каждый из вопросов;
- уясните, какие учебные элементы остались для вас неясными и постарайтесь получить на них ответ заранее (до семинарского занятия) во время текущих консультаций преподавателя;
- готовиться можно индивидуально, парами или в составе малой группы, последние являются эффективными формами работы;
- рабочая программа дисциплины в части целей, перечню знаний, умений, терминов и учебных вопросов может быть использована вами в качестве ориентира в организации обучения.

Выполнение задания:

- выберите набор данных (временной ряд, временные ряды) для выполнения задания;
- выполните анализ используемых признаков (целевого признака);
- проанализируйте качество исходных данных;
- выполните выбор инструментов предобработки для улучшения качества исходных данных, а также формулировки предварительных гипотез;
- решите задачу прогнозирования уровней временного ряда;
- исследуйте возможность извлечения признаков временного ряда;
- решите задачу анализа выявленных признаков;
- оформите отчет по результатам выполнения задания.

7. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

7.1. Основная литература

1. Афанасьев, Владимир Николаевич. Анализ временных рядов и прогнозирование. - Саратов: Ай Пи Ар Медиа. – 310 с. Текст : электронный. - URL: <http://www.iprbookshop.ru/90196.html> (дата обращения: 12.11.2020). - Режим доступа: для авторизир. пользователей
2. Голоктионова Ю.Г., Ильминская С.А., Илюхина И.Б., Луговской А.М., Лисичкина Н.В. и др. Прогнозирование и планирование в экономике. - Москва: Прометей –

- 544 с. Текст : электронный. - URL: <http://www.iprbookshop.ru/94511.html> (дата обращения: 01.10.2020). - Режим доступа: для авторизир. пользователей
3. Мاستицкий С. Э. (2020) Анализ временных рядов с помощью R. — Электронная книга, адрес доступа: <https://ranalytics.github.io/tsa-with-r>
 4. Миркин, Борис Григорьевич. Введение в анализ данных – М.:Юрайт, 2020 – 174 с. Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262> (дата обращения: 01.10.2020)
 5. Мхитарян В. С., Архипова М. Ю., Дуброва Т. А., Миронкина Ю. Н., Сиротин В. П. Анализ данных. – М.: Юрайт, 2020 – 490 с. Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166> (дата обращения: 29.09.2020)
 6. О'Нил, Кэти. Data Science : Инсайдерская информация для новичков. Включая язык R : [пер. с англ.] – СПб. Питер. – 368 с. Текст: электронный. - URL: <http://new.ibooks.ru/bookshelf/359209/reading> (дата обращения: 25.01.2021)
 7. Хайндман Р. Дж, Атанасопулос Дж. Прогнозирование: принципы и практика. [Электронный ресурс] –URL: <https://otexts.com/fpp3/>

Все источники основной литературы взаимозаменяемы.

7.2. Дополнительная литература

8. Барсегян А.А, Куприянов М.С., Степаненко В.В., Холод И.И. Прогнозирование временных рядов и процессов. 3-е изд. - СПб. : БХВ-Петербург, 2010. - 512 с.
9. Жерон, Орельен. ПрМ.:Диалектика. -684 с. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow : Концепции, инструменты и техники для создания интеллектуальных систем : полноцветное издание : перевод с английского -
10. Principles of Econometrics with R [Электронный ресурс] – URL: <https://bookdown.org/ccolonescu/RPoE4/>
11. Нильсен Эйлин. Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение. –М.: ООО Диалектика – 2021 – 544 с.
12. Наумов, Владимир Николаевич. Средства бизнес- аналитики: учеб. пособие / В. Н. Наумов ; Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2016. - 107 с.
13. Наумов В.Н. Анализ данных и машинное обучение: мтеоды и инструментальные средства. Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2020. - 260 с.

7.3. Нормативные правовые документы и иная правовая информация

Не используются

7.4. Интернет-ресурсы.

СЗИУ располагает доступом через сайт научной библиотеки <http://nwapr.spb.ru/> к следующим подписным электронным ресурсам:

<https://ranalytics.github.io/tsa-with-r/ch-intro-to-prophet.html>

Русскоязычные ресурсы

Электронные учебники электронно - библиотечной системы (ЭБС) «Айбукс»

Электронные учебники электронно – библиотечной системы (ЭБС) «Лань»

Рекомендуется использовать следующий интернет-ресурсы

<http://serg.fedosin.ru/ts.htm>

<http://window.edu.ru/resource/188/64188/files/chernyshov.pdf>

7.5. Иные источники.

Не используются.

8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Учебная дисциплина включает использование программного обеспечения Microsoft Excel, Microsoft Word, для подготовки текстового и табличного материала.

Интернет-сервисы и электронные ресурсы (поисковые системы, электронная почта, профессиональные тематические чаты и форумы, системы аудио и видео конференций, онлайн энциклопедии, справочники, библиотеки, электронные учебные и учебно-методические материалы).

Для организации дистанционного обучения используется система Moodle.

№ п/п	Наименование
1.	Компьютерные классы с персональными ЭВМ, объединенными в локальные сети с выходом в Интернет
2.	Пакет Excel -2013, 2016, professional plus
3.	Аналитическая платформа Qlik View, MS BI
4.	Система бизнес-аналитики Deductor Academic, Loginom
5.	Средства интеллектуального анализа SQ Lserver. Настройка Analysis services, data mining ad-insfor Office.
6.	Облачные технологии Loginom –server, Google colab
7.	Язык R, Python, Anaconda navigator, Rstudio
8.	Мультимедийные средства в каждом компьютерном классе и в лекционной аудитории
9.	Браузер, сетевые коммуникационные средства для выхода в Интернет
10.	Система дистанционного обучения Moodle

Компьютерные классы из расчета 1 ПЭВМ для одного обучаемого. Каждому обучающемуся должна быть предоставлена возможность доступа к сетям типа Интернет в течение не менее 20% времени, отведенного на самостоятельную подготовку.