

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Национальный исследовательский
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ:

Декан ЭФ

В.Е. Сахарова

2014 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ЭКОНОМИЧЕСКАЯ ИНФОРМАТИКА

Направление подготовки

38.06.01 «Экономика»

Квалификация (степень) выпускника

Исследователь, преподаватель-исследователь

Форма обучения

Очная (Заочная)

Профили подготовки:

08.00.01 – Экономическая теория

08.00.05 – Экономика и управление народным хозяйством (экономика труда)

08.00.10 – Финансы, денежное обращение и кредит

Статус дисциплины: Блок 1 «Образовательные дисциплины»

Вариативная часть

Программа одобрена на заседании учебно-методической
комиссии экономического факультета
Томского государственного университета
от « 2 » 09 2014 года, протокол № 1

г. Томск – 2014

Рабочая программа составлена на основании федеральных государственных образовательных стандартов к основной образовательной программе высшего образования подготовки научно-педагогических кадров в аспирантуре по направлению 38.06.01 «Экономика».

Автор-разработчик:

- ❖ *Богданов Александр Леонидович*, кандидат технических наук, доцент кафедры математических методов и информационных технологий в экономике

АННОТАЦИЯ ПРОГРАММЫ

Дисциплина «Экономическая информатика» принадлежит к Блоку 1 «Образовательные дисциплины», и относится к обязательным дисциплинам вариативной части учебного плана (Б1.В.ОД.6). Она содержательно знакомит слушателей с современными математическими методами обработки и анализа массивов экономических данных, накапливаемых в корпоративных информационных системах, с целью извлечения скрытых знаний и проверки гипотез; формирование навыков практической работы с современными компьютерными пакетами прикладных программ анализа и поиска скрытых закономерностей в "сырых" данных.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Экономическая информатика» является знакомство с современными математическими методами обработки и анализа массивов экономических данных, накапливаемых в корпоративных информационных системах, с целью извлечения скрытых знаний и проверки гипотез; формирование навыков практической работы с современными компьютерными пакетами прикладных программ анализа и поиска скрытых закономерностей в "сырых" данных.

Изучение дисциплины предполагает выполнение следующих задач:

- ❖ совершенствование умений обучающихся (аспирантов) в выборе методов, способов получения, хранения, переработки информации и навыков работы с компьютером как средством управления информацией;
- ❖ совершенствование умений в использовании новейших информационно-коммуникационных технологий и организации информационного поиска в глобальных компьютерных сетях;
- ❖ знакомство с современными технологиями электронного обучения.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП АСПИРАНТА

Дисциплина «Экономическая информатика» принадлежит к Блоку 1 «Образовательные дисциплины», и относится к обязательным дисциплинам вариативной части учебного плана (Б1.В.ОД.6). Она содержательно знакомит слушателей с современными математическими методами обработки и анализа массивов экономических данных, накапливаемых в корпоративных информационных системах, с целью извлечения скрытых знаний и проверки гипотез; формирование навыков практической работы с современными компьютерными пакетами прикладных программ анализа и поиска скрытых закономерностей в "сырых" данных.

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

3.1 Компетенции, формируемые в результате освоения дисциплины:

- ❖ способностью самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием

современных методов исследования и информационно-коммуникационных технологий (ОПК-1);

- ❖ способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленности "Экономическая теория" (ПК-1);
- ❖ способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленности "Экономика и управление народным хозяйством: экономика труда" (ПК-2);
- ❖ способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленности "Финансы, денежное обращение и кредит" (ПК-3).

3.2 Требования к результатам освоения содержания дисциплины:

В результате обучения аспирант должен:

ЗНАТЬ

- ❖ современные математические методы анализа и поиска скрытых закономерностей в больших массивах экономических данных, накапливаемых в корпоративных информационных системах; знать условия их применимости;

УМЕТЬ

- ❖ применять математические методы анализа и поиска скрытых закономерностей в больших массивах экономических данных для решения практических задач и критически оценивать получаемые результаты;

ВЛАДЕТЬ

- ❖ навыками работы с современными компьютерными пакетами прикладных программ анализа и поиска скрытых закономерностей в "сырых" данных.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 1 зачетная единица, 36 часов.

№ п/п	Раздел дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)			Формы текущего контроля успеваемости Форма промежуточной аттестации
				лекц ии	практи- ческие занятия	самосто- ятельная работа	
1	Методы и инструменты трансформации данных	4	1	1	0	2	Практическое задание. Контрольная работа.
2	Методы и инструменты очистки данных	4	2	1	0	2	Практическое задание. Контрольная работа.
3	Инструменты визуализации данных	4	3	1	0	2	Практическое задание. Контрольная работа.
4	Задача ассоциации	4	4	1	2	2	Практическое задание. Контрольная работа.
5	Задача кластеризации	4	6	1	2	2	Практическое задание. Контрольная работа.

6	Задача классификации и регрессии	4	8	1	4	4	Практическое задание. Контрольная работа.
7	Зада анализа и прогнозирования временных рядов	4	14	0	2	2	Практическое задание. Контрольная работа.
8	Ансамблимоделей	4	16	0	2	2	Практическое задание. Контрольная работа.
	Всего			6	12	18	

Самостоятельная работа аспирантов по темам дисциплины заключается в выполнении практических заданий, в том числе в системе электронного обучения. При этом самостоятельная работа разделяется на самостоятельную работу по освоению теоретического материала, а также выполнение практических заданий и работу по представлению результатов создания персональных текстов.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В учебном процессе используются: лекции, практические занятия, внеаудиторная работа с целью формирования и развития профессиональных навыков обучающихся.

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В процессе преподавания дисциплины «Академическое письмо» применяются методы текущего и итогового контроля знаний.

6.1 Контрольные вопросы и задания для проведения текущего контроля

1) по разделу «Методы и инструменты трансформации данных»

Назначение инструментов трансформации данных. Методы трансформации упорядоченных данных. Методы группировки данных. Методы слияния данных. Методы квантования, нормализации и кодирования данных.

2) по разделу «Методы и инструменты очистки данных»

Назначение инструментов очистки данных. Методы оценки качества данных. Очистка, предобработка и фильтрация данных. Обработка дубликатов и противоречий. Обнаружение аномальных значений. Восстановление пропущенных значений. Задачи сокращения размерности, сокращение числа признаков и числа значений признаков. Сэмплинг.

3) по разделу «Инструменты визуализации данных»

Цели и задачи инструментов визуализации данных. Базовые визуализаторы. Визуализаторы многомерных данных. Визуализаторы для оценки качества моделей и интерпретации результатов.

4) по разделу «Задача ассоциации»

Постановка задачи выявления ассоциаций. Алгоритм Apriori. Иерархические ассоциативные правила. Поиск ассоциативных правил с учётом фактора времени.

5) по разделу «Задача кластеризации»

Постановка задачи кластеризации. Агломеративные алгоритмы кластеризации. Алгоритмы k-средних и g-средних. Сети и карты Кохонена.

6) по разделу «Задача классификации и регрессии»

Постановка задачи классификации. Модель множественной линейной регрессии. Оценка качества модели множественной линейной регрессии. Модели регрессии с качественными переменными. Деревья решений. Нейронные сети.

7) по разделу «Задача анализа и прогнозирования временных рядов»

Определение понятия «временной ряд». Цели исследования и области практического применения временных рядов. Компоненты временных рядов: систематическая, сезонная, циклическая и случайная. Стационарные временные ряды.

8) по разделу «Ансамбли моделей»

Понятия ансамбля моделей. Бэггинг и Бустинг. Методы оценки эффективности и сравнения моделей.

6.2 Контрольные вопросы для промежуточной аттестации (вопросы к экзамену)

Тема 1. Обзор технологий анализа данных

1. Что входит в понятие «анализ данных»?
2. Перечислите основные подходы к моделированию.
3. Что такое «адекватность» модели?
4. Почему для информационного подхода к моделированию требуется высокое качество исходных данных?
5. Назовите основные преимущества информационного подхода к моделированию перед другими подходами.
6. Из каких основных этапов состоит процесс построения моделей?
7. Что такое транзакция?
8. Перечислите основные методы сбора данных для анализа, а также их основные достоинства и недостатки.
9. Что означают следующие понятия: KDD и Data Mining?
10. Что такое трансформация данных? В чём заключается её отличие от предобработки и очистки?
11. Перечислите основные причины, по которым требуется осуществлять трансформацию данных в аналитическом приложении.
12. Что такое временной ряд? Приведите примеры.
13. Опишите принцип работы механизма преобразования данных скользящим окном.
14. Дайте определение понятиям «глубина погружения» и «горизонт прогноза».
15. Опишите основные методы трансформации дат.
16. Что такое группировка данных?
17. Перечислите основные причины группировки данных?
18. Перечислите основные функции агрегации.
19. Что такое соединение данных?
20. Перечислите основные методы соединения данных.
21. Что такое квантование данных?
22. Опишите основные подходы к выполнению квантования данных.
23. Что такое операция транспонирования?
24. Что такое обратное транспонирование?
25. Что такое нормализация данных?
26. Перечислите основные причины, по которым требуется выполнять нормализацию данных.
27. Перечислите основные методы нормализации данных.
28. Как осуществляется нормализация категориальных признаков?
29. Что понимается под термином «очистка данных»?
30. Что такое «качество данных»?
31. Перечислите основные методы оценки качества данных.
32. Как влияет на качество данных методика их сбора и метод их хранения?
33. Что такое «профайлинг данных»?
34. Какие проблемы можно обнаружить, осуществляя визуальный анализ таблиц с

данными?

35. Какие проблемы можно обнаружить, осуществляя визуальный анализ графиков и диаграмм данных?
36. Какие ошибки в данных являются трудноформализуемыми?
37. Почему очистка данных может потребоваться даже после прохождения всех процедур ETL-процессов?
38. Что такое мониторинг качества данных?
39. Почему некоторые проблемы качества данных приходится решать в процессе их предобработки в аналитическом приложении?
40. Что такое снижение размерности входных данных и устранение незначущих признаков?
41. Почему окончательная предобработка данных может быть выполнена только в аналитическом приложении и только с учетом требований конкретной задачи анализа?
42. Что такое фильтрация данных?
43. Что такое дубликаты и противоречия?
44. В каких случаях дубликаты и противоречия не являются следствием ошибок?
45. Какие методы используются для устранения дубликатов и противоречий?
46. Что такое аномальные значения и какие методы используются для их обнаружения и исправления?
47. Что такое пропуски в данных?
48. Перечислите основные причины появления пропусков в данных.
49. Перечислите основные методы устранения пропусков в данных.
50. В чем заключается различие подходов восстановления пропусков в упорядоченных и неупорядоченных данных?
51. Что такое снижение размерности исходных данных?
52. Какие цели преследуются при проведении сокращения размерности данных?
53. Перечислите основные методы сокращения размерности.
54. Чем отличаются незначущие признаки от избыточных?
55. Что такое энтропия?
56. Что такое метод главных компонент?
57. Что такое сокращение числа значений признаков и записей в исходной выборке данных?
58. Что такое сэмплинг?
59. Перечислите основные методы сэмплинга?
60. Каковы цели и задачи визуализации данных в аналитических технологиях?
61. Почему требуется визуальный контроль данных в аналитическом приложении?
62. Какие инструменты визуализации данных относят к визуализаторам общего назначения?
63. Перечислите основные виды диаграмм.
64. Что такое гистограмма?
65. Как работает визуализатор OLAP-кубов?
66. Что такое кросс-диаграмма?
67. Что такое таблица сопряженности?
68. О чём говорят значения, расположенные на главной диагонали таблицы сопряженности?
69. Что такое диаграмма рассеяния?
70. Каким образом можно применить диаграмму рассеяния для анализа качества модели?
71. В каких целях визуализируют выходную ошибку модели?
72. Дайте определение понятия «ассоциация».
73. Что понимается под транзакцией в теории ассоциативных правил?
74. В чём заключается постановка задачи анализа рыночной корзины?

75. Что такое «поддержка» и «достоверность»?
76. Что такое «значимость» и «полезность» ассоциативных правил?
77. В чём состоит основная проблема поиска ассоциативных правил?
78. Покажите, что стратегия поиска ассоциативных правил основанная на проверке поддержки и достоверности всех возможных ассоциаций неэффективна.
79. Что такое «частота предметного набора»?
80. Какие предметные наборы называются популярными (или частыми)?
81. Дайте описание алгоритма apriori.
82. Что такое «иерархия предметов» в задачах поиска ассоциативных правил?
83. Какие имеются достоинства и недостатки подхода поиска ассоциативных правил с учетом иерархии предметов?
84. На каких уровнях иерархии чаще всего встречаются предметы, ассоциации между которыми имеют большую поддержку?
85. Почему, для объекта, редко встречающегося на верхних уровнях иерархии, анализировать ассоциации с его потомками на нижних уровнях не имеет смысла?
86. С какой целью применяют понижение порога минимальной поддержки для нижних уровней иерархии?
87. В чём заключается постановка задачи кластеризации?
88. Что такое метрика?
89. Опишите алгоритм k-means?
90. Что такое центроид, как он рассчитывается и какова его роль в алгоритме k-means?
91. Перечислите возможные условия останова алгоритма k-means.
92. Каким образом в алгоритме k-means определяется принадлежность точки к кластеру?
93. Перечислите достоинства и недостатки алгоритма k-means?
94. Что такое сеть Кохонена?
95. Опишите процедуру конкуренции, объединения и подстройки весов нейронов сети Кохонена.
96. Что такое радиус обучения в алгоритме обучения сети Кохонена?
97. В чём заключается принцип самоорганизации?
98. Какие факторы влияют на выбор числа выходных нейронов сети Кохонена?
99. Что такое принцип сохранения топологического подобия?
100. Что такое карта Кохонена и чем она отличается от сети Кохонена.
101. Каким образом раскраска карты Кохонена позволяет судить о качестве кластеризации?
102. Что такое «мертвые» нейроны и почему они появляются?
103. Перечислите критерии, влияющие на выбор числа нейронов (ячеек) карты Кохонена?
104. Опишите основную идею агломеративных алгоритмов кластеризации.
105. Кто такое дендрограмма и график роста межкластерных расстояний?
106. Перечислите основные проблемы применения алгоритмов кластеризации.
107. Что такое множественная линейная регрессия?
108. Что такое остатки в модели линейной регрессии?
109. Дайте интерпретацию коэффициентам модели множественной линейной регрессии.
110. Что такое стандартная ошибка оценивания?
111. Что такое коэффициент детерминации?
112. Перечислите предпосылки, лежащие в основе метода наименьших квадратов.
113. Что означает «значимость» регрессионной модели?
114. Какие статистики используются и какие статистические гипотезы тестируются при проверке значимости регрессионной модели?
115. Какие особенности следует учитывать при оценивании качества моделей множественной линейной регрессии по значению коэффициента детерминации?

116. Что такое «фиктивные переменные»?
117. Что такое опорная категория?
118. Дайте интерпретацию коэффициентам модели множественной линейной регрессии, построенной только на фиктивных переменных.
119. Что такое логистическая регрессия?
120. Приведите примеры бизнес-задач, решение которых требует применения моделей с бинарной выходной переменной.
121. Как интерпретируются результаты логистической регрессии?
122. Что такое отношение правдоподобия?
123. Что такое доверительный интервал?
124. Какие тесты используются для проверки гипотез о значимости одного или группы коэффициентов?
125. Каким образом можно использовать логистическую регрессию для решения задач классификации?
126. В чем заключается задача отбора входных переменных для построения регрессионной модели?
127. Какой метод отбора переменных приводит к наилучшему набору объясняющих факторов?
128. Что такое мультиколлинеарность?
129. Каким образом можно устранить мультиколлинеарность?
130. Что такое гетероскедастичность?
131. Что такое деревья решений?
132. Опишите алгоритм построения дерева решений ID3.
133. Опишите алгоритм построения дерева решений C4.5
134. Опишите алгоритм построения дерева решений CART.
135. С какой целью осуществляется упрощение деревьев решений.
136. Что такое нейронные сети.
137. Опишите модель искусственного нейрона.
138. Перечислите несколько основных подходов построения нейронных сетей.
139. Опишите процедуру обучения нейронных сетей.
140. Опишите алгоритм обратного распространения ошибки.
141. Определение понятия «временной ряд».
142. Перечислите цели исследования и области практического применения временных рядов.
143. Перечислите основные компоненты временных рядов.
144. Дайте определение стационарного временного ряда.
145. Что такое эргодический процесс?
146. Перечислите основные модели однофакторных моделей временных рядов.
147. Какие информационные критерии используются для оценивания качества идентификации моделей временных рядов?
148. Каким образом производится прогнозирование на основе моделей временных рядов?
149. Опишите подходы к оцениванию эффективности и сравнению моделей.
150. К чему может привести ошибочная классификация?
151. Что такое Lift- и Profit-кривые?
152. В чем заключается основная идея ROC-анализа?
153. Что такое ансамбли моделей?
154. Опишите основные идеи, лежащие в основе бэггинга и бустинга.

7.1. Основная литература

1. Архипников С.Я., Голубев Д.В., Максименко О.Б. Хранилища дан-ных. – М.: Диалог-МИФИ, 2002.
2. Барсегян А.А. и др. Технологии анализа данных: Data Mining, Visual Mining, OLAP. 2-ое изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007.
3. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Но-восибирск: Изд-во Ин-та математики, 1999.
4. Зиновьев А.Ю. Визуализация многомерных данных. – Красноярск: Изд-во КГТУ, 2000.
5. Паклин Н.Б., Орешков В.И., Бизнес-аналитика: от данных к знаниям: Учеб. Пособие. 2-е издание, перераб. и доп. – СПб.:Питер 2010.
6. Спирли Э. Корпоративные хранилища данных. Планирование, разра-ботка, реализа-ция. – М.: Вильямс, 2001. Т. 1.
7. Хайкин С. Нейронные сети: полный курс. 2-е изд. – М.: Вильямс, 2006.

7.2. Дополнительная литература

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконо-метрики. М.: ЮНИТИ, 1998.
2. Андерсон Т. Статистический анализ временных рядов. М.: Мир, 1976.
3. Доугерти К. Введение в эконометрику. М.: ИНФРА-М, 1997.
4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Финан-сы и статистика, 1986.
5. Канторович Г.Г. Лекции по курсу «Анализ временных рядов». Эко-номический жур-нал ВШЭ, 2002.
6. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976.
7. Кремер Н.Ш., Путко Б.А. Эконометрика. М., ЮНИТИ, 2002.
8. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. На-чальный курс. М.: Дело, 2000.

7.3. Периодические издания и источники в сети INTERNET

1. Журнал «Открытые системы», <http://www.osp.ru/os/index.html>
2. Информационный бюллетень Jet Info Online, <http://www.jetinfo.ru/>
3. Журнал «Windows IT Pro», <http://www.osp.ru/win2000/index.html>
4. Сервер «CIT Forum», <http://citforum.ru/>
5. Журнал «КомпьютерПресс», <http://www.compress.ru/>

7.4. Программное и коммуникационное обеспечение

1. Deductor, разработчик BaseGroup Labs, <http://www.basegroup.ru/>
2. EViews, разработчик: Quantitative Micro Software, <http://www.eviews.com/>
3. R – язык программирования и среда для статистической обработки данных и работы с графикой, разработчик: R Foundation, <http://www.r-project.org/>
4. RapidMiner, разработчик Rapid-I GmbH, <http://www.rapidminer.com/>

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Лекционная аудитория, аудитория для проведения практических занятий и консульта-ций, доступ к сети Интернет.