

СОДЕРЖАНИЕ

Введение.....	5
1. Построение деревьев решений – система See5/C5.0	6
1.1. Подготовка данных для See5.....	6
Файл имен переменных.....	7
Файл данных	8
Файлы тестовых данных (необязательные)	9
Файл стоимости	9
1.2. Интерфейс пользователя.....	9
1.3. Усиление решения (Boosting).....	13
1.4. Использование правил для принятия решений	14
1.5. Смягчение порогов.....	15
1.6. Дополнительные настройки алгоритма.....	15
1.7. Перекрестная проверка	15
1.8. Выборка из больших наборов данных	16
1.9. Учет стоимости различных ошибок классификации	17
1.10. Использование классификаторов.....	18
1.11. Детальная проверка и сохранение результатов	18
2. WizWhy – система поиска логических правил в данных.....	20
2.1. Общие свойства системы WizWhy	20
2.2. Загрузка и управление данными	21
2.3. Задание параметров процедуры поиска правил	23
2.4. Работа с окном диалога Ошибки/Примеры (Errors/Examples).....	25
2.5. Работа с другими окнами диалога	25
2.6. Результаты работы системы	26
2.7. Предсказание на основе полученных правил	36
3. Практические примеры	40
3.1. Сравнение структуры интеллекта физиков и лириков	40
Общая характеристика данных	40
Сравнение средних значений результатов тестирования в группах физиков и лириков	42
Поиск логических закономерностей системой WizWhy	44
3.2. Влияние возраста и стажа работников на производительность труда.....	47
Дисперсионный анализ	48
Обработка данных системой WizWhy	53
3.3. Выяснение причин неурожайности сельскохозяйственных участков	55
Исходные данные	55
Комплексная обработка данных традиционными методами	56
Результаты обработки данных системой See5	59
Результаты обработки данных системой WizWhy	60
3.4. Прогнозирование продолжительности ремиссий при алкоголизме.....	62

Общая характеристика данных	63
Частотный анализ признаков.....	66
Дискриминантный анализ.....	67
Результаты обработки данных системой WizWhy	68
Результаты обработки данных системой See5 (decision trees)	72
Отчет системы See5	73
Заключение	75
Литература	77