

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**Пермский национальный исследовательский
политехнический университет**

Строительный факультет

Кафедра Строительный инжиниринг и материаловедение

Практическое задание

по программе повышения квалификации «Интеллектуальный анализ
данных с помощью методов системно-когнитивного анализа»

на тему:

Определение статуса банкрота среди застройщиков
Пермского края на основе АСК-анализа

Выполнила:
студентка гр. ЭУН-17-16
Чугайнова А.М

Руководитель работы:
канд. экон. наук, доцент
Алексеев А.О.

Пермь 2021 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1.СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	4
1.1. Описание решения	4
1.2. Преобразование исходных данных в файл MS Excel.....	4
1.3. Синтез и верификация статистических и интеллектуальных моделей	9
1.4. Виды моделей системы «Эйдос».....	10
1.5. Результаты верификации моделей	12
2.РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	14
2.1. Решение задачи идентификации.....	14
2.2. Когнитивные функции	17
2.3. SWOT и PEST матрицы и диаграммы	19
2.4. Нелокальные нейронные сети	22
2.5. Паретто-подмножества нелокальной нейронной сети.....	22
2.6. Когнитивные диаграммы	24
ЗАКЛЮЧЕНИЕ	27
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	28

ВВЕДЕНИЕ

Одним из важнейших и перспективных направлений в развитии современных информационных технологий, на сегодняшний день, является создание систем искусственного интеллекта. В результате развития технологий в данном направлении, возникла необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи идентификации статусов застройщиков.

При выполнении работы нужно учесть, что необходимы свободный доступ к исходным данным и методика, которая необходима для преобразования этих данных в форму, которая необходима для работы в системе искусственного интеллекта. Среди баз данных для систем искусственного интеллекта удачным выбором являются: единая информационная система жилищного строительства (<https://наш.дом.рф>), единый ресурс застройщиков (<https://erzrf.ru>), каталог организаций (<https://www.list-org.com>).

Для решения задачи используем стандартные возможности Microsoft Office Excel, а также систему искусственного интеллекта «Aidos – X ++».

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

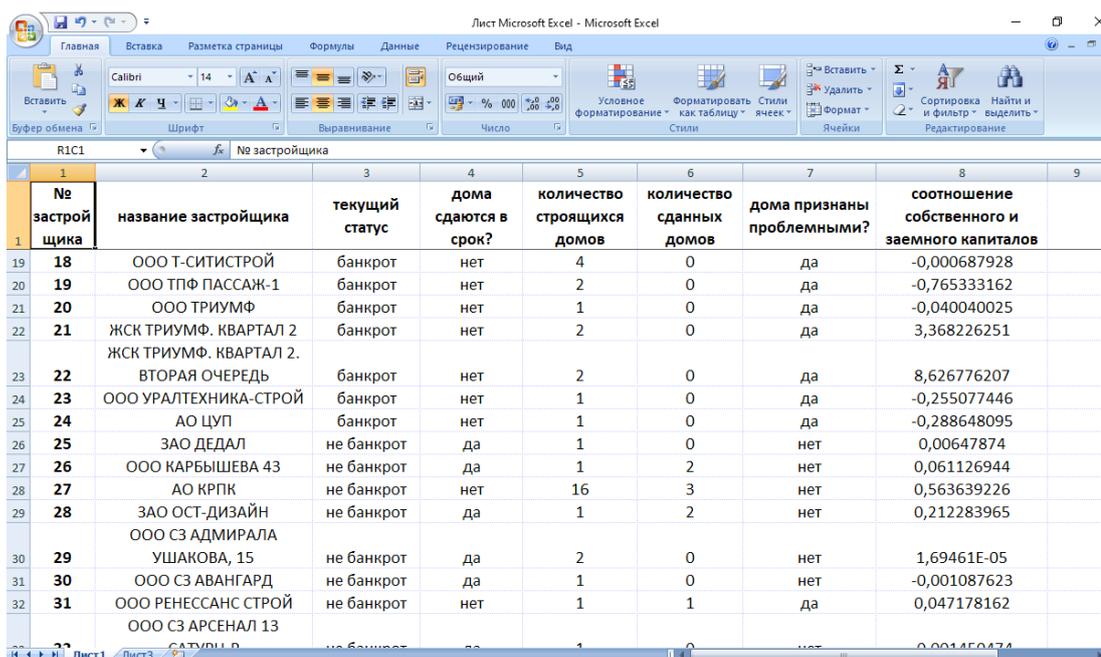
1.1. Описание решения

В соответствии с методологией АСК-анализа решение поставленной задачи проведем в четыре этапа:

1. Преобразование исходных данных в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы «Эйдос».
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2. Преобразование исходных данных в файл MS Excel

Из электронных ресурсов создали базу данных «Застройщики Пермского края», в которую вошло 24 застройщика-банкрота и 24 «здоровых» застройщика. Исходные данные представлены на рисунке 1.



№ застройщика	название застройщика	текущий статус	дома сдаются в срок?	количество строящихся домов	количество сданных домов	дома признаны проблемными?	соотношение собственного и заемного капиталов
18	ООО Т-СИТИСТРОЙ	банкрот	нет	4	0	да	-0,000687928
19	ООО ТПФ ПАССАЖ-1	банкрот	нет	2	0	да	-0,765333162
20	ООО ТРИУМФ	банкрот	нет	1	0	да	-0,040040025
21	ЖСК ТРИУМФ. КВАРТАЛ 2 ЖСК ТРИУМФ. КВАРТАЛ 2.	банкрот	нет	2	0	да	3,368226251
22	ВТОРАЯ ОЧЕРЕДЬ	банкрот	нет	2	0	да	8,626776207
23	ООО УРАЛТЕХНИКА-СТРОЙ	банкрот	нет	1	0	да	-0,255077446
24	АО ЦУП	банкрот	нет	1	0	да	-0,288648095
25	ЗАО ДЕДАЛ	не банкрот	да	1	0	нет	0,00647874
26	ООО КАРБЫШЕВА 43	не банкрот	да	1	2	нет	0,061126944
27	АО КРПК	не банкрот	нет	16	3	нет	0,563639226
28	ЗАО ОСТ-ДИЗАЙН	не банкрот	да	1	2	нет	0,212283965
29	ООО СЗ АДМИРАЛА УШАКОВА, 15	не банкрот	да	2	0	нет	1,69461E-05
30	ООО СЗ АВАНГАРД	не банкрот	да	1	0	нет	-0,001087623
31	ООО РЕНЕССАНС СТРОЙ ООО СЗ АРСЕНАЛ 13	не банкрот	нет	1	1	да	0,047178162
32	ООО СЗ АРСЕНАЛ 13	не банкрот	нет	1	0	нет	0,001450474

Рисунок 1 - Исходные данные в файле MS Excel

Поскольку ввод исходных данных в систему «Эйдос» планируется осуществить с помощью ее универсального программного интерфейса импорта данных из внешних баз данных, который работает с файлами MS Excel, то преобразуем данные в xls-файл, для чего выполним следующие операции.

Скопируем таблицу в MS Excel и запишем ее с именем: Inp_data.xls в папку: C:\Aidos-X\AID_DATA\Inp_data\. В файле Inp_data.xls добавим пустую колонку на позиции «А» и автоматически пронумеруем все строки. В результате получим таблицу исходных данных, полностью подготовленную для обработки в системе «Эйдос» и записанную в нужную папку в виде файла нужного типа с нужным именем.

Для загрузки базы исходных данных в систему «Эйдос» необходимо воспользоваться универсальным программным интерфейсом для ввода данных из внешних баз данных табличного вида, расположенный в режиме 2.3.2.2.

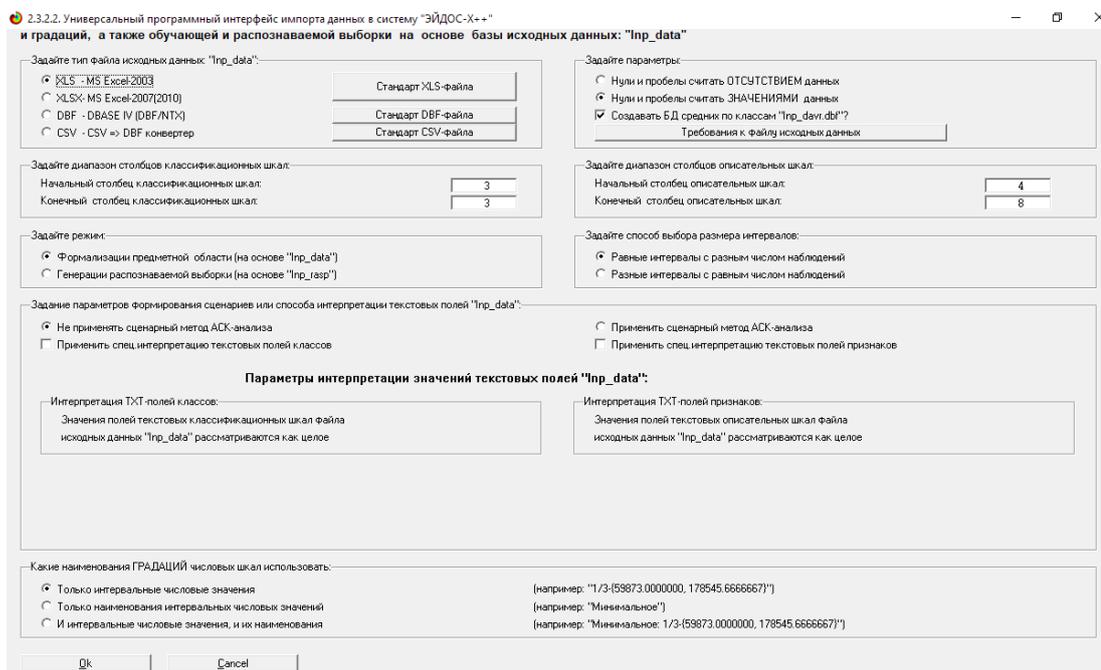


Рисунок 2- Экранная форма универсального программного интерфейса импорта данных в систему «Эйдос»

В экранной форме, приведенной на рисунке 2, необходимо задать настройки:

- «Тип файла исходных данных Inp_data»: «XLS – MS Excel-2003»;
- «Диапазон столбцов классификационных шкал»: «Начальный столбец классификационных шкал» – 3, «Конечный столбец классификационных шкал» – 3;
- «Диапазон столбцов описательных шкал»: «Начальный столбец описательных шкал»-4, «Конечный столбец описательных шкал»-8;
- «Параметры формирования сценариев или способа интерпретации текстовых полей»: «Не применять сценарный метод АСК-анализа».

После нажать кнопку «ОК». После нажатия кнопки открывается окно, где размещена информация о размерности модели. В этом окне необходимо нажать кнопку «Выйти на создание модели», что показано на рисунке 3.

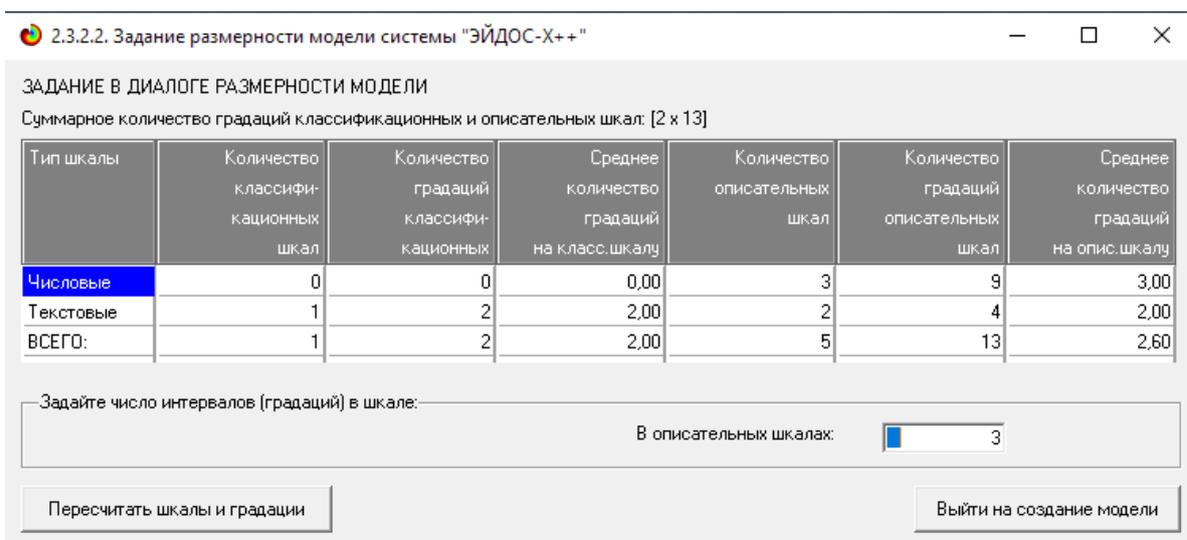


Рисунок 3- Задание размерности модели системы «Эйдос»

Далее следует открытие окна, которое отображает стадию процесса импорта данных из внешней БД «Inp_data.xls» в систему «Эйдос», а также прогноз времени завершения этого процесса. В данном окне необходимо дождаться завершения формализации предметной области и нажать кнопку «ОК», как показано на рисунке 4.

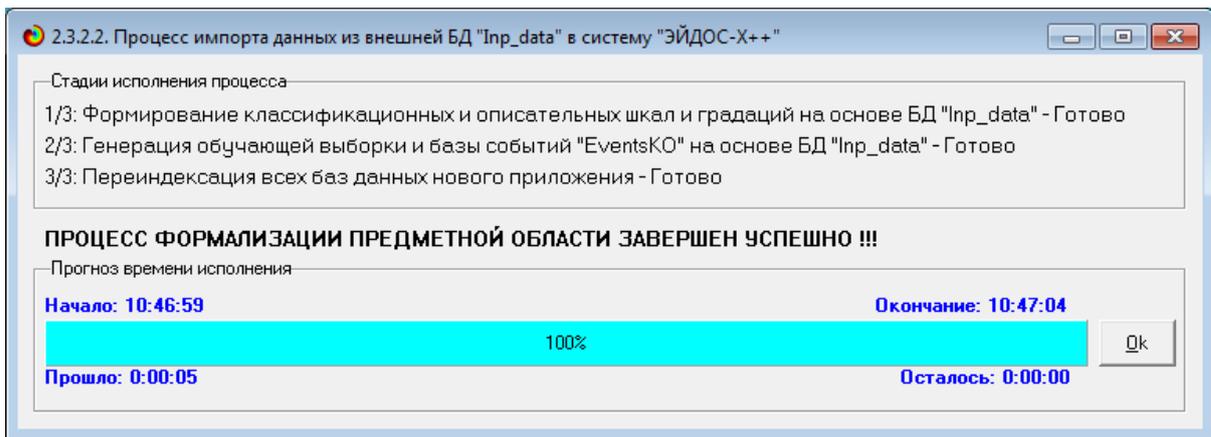


Рисунок 4- Процесс импорта данных из внешней БД «Inp_data.xls» в систему «Эйдос»

В результате выполненных действий формируются классификационные и описательные шкалы, градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Эти действия полностью автоматизированы и выполняются на 2-м этапе АСК – анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1, рисунок 5.

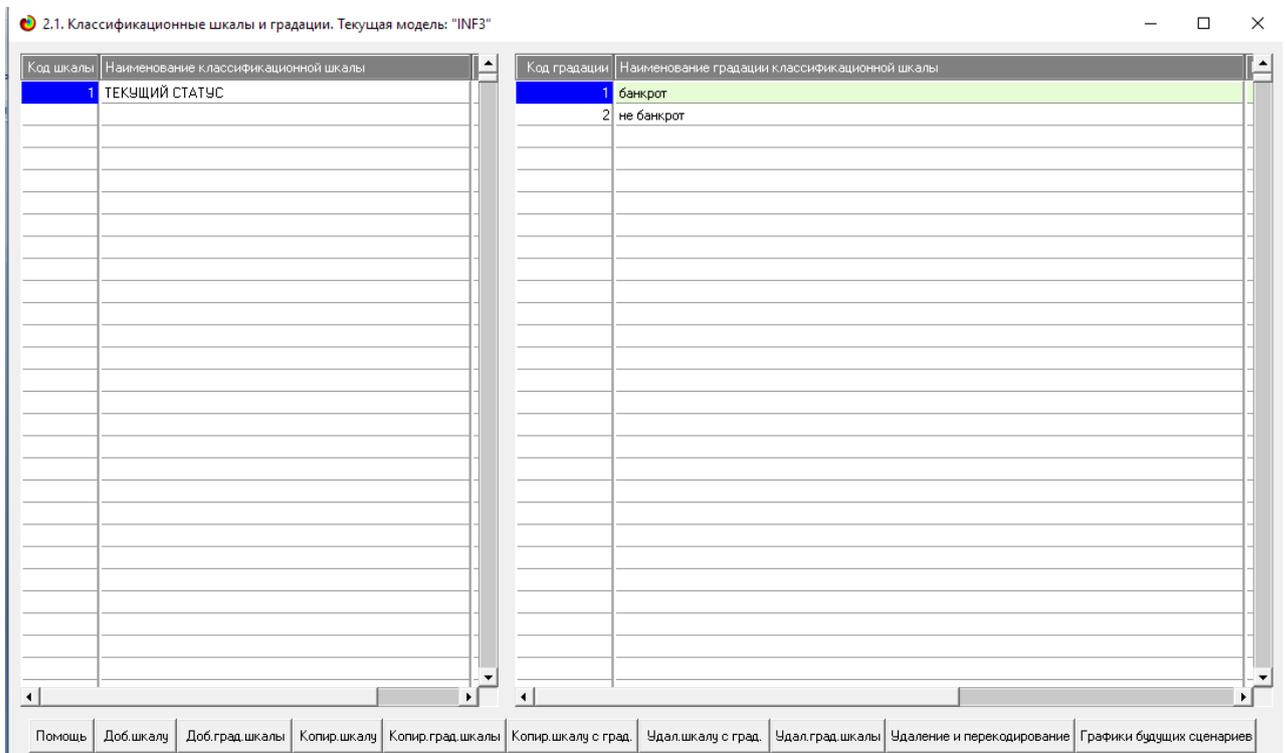


Рисунок 5- Классификационные шкалы и градации (фрагмент)

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2, рисунок 6.

Код шкалы	Наименование описательной шкалы	Код градации	Наименование градации описательной шкалы
1	ДОМА СДАЮТСЯ В СРОК?	11	1/3-(-0.7895236, 0.3492430)
2	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ	12	2/3-(-0.3492430, 2.4880096)
3	КОЛИЧЕСТВО СДАННЫХ ДОМОВ	13	3/3-(2.4880096, 8.6267762)
4	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ?		
5	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТА...		

Рисунок 6- Описательные шкалы и градации (фрагмент)

Для просмотра обучающей выборки необходимо запустить режим 2.4., рисунок 7.

№	Наименование объекта	3. ТЕКУЩИЙ СТАТУС	4. ДОМА СДАЮТСЯ В СРОК?	5. КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ	6. КОЛИЧЕСТВО СДАННЫХ ДОМОВ	7. ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ?	8. СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ
1	1	1	2	3	6	9	11
2	2	1	2	3	6	9	11
3	3	1	2	3	6	9	11
4	4	1	2	3	6	9	11
5	5	1	2	3	6	9	11
6	6	1	2	5	6	9	11
7	7	1	2	3	6	9	11
8	8	1	2	3	6	9	11
9	9	1	2	3	6	9	11
10	10	1	2	3	6	9	11
11	11	1	2	3	6	9	11
12	12	1	2	3	6	9	11
13	13	1	2	3	6	9	11
14	14	1	2	3	6	9	11
0	1	3	4	5	6	7	8

Рисунок 7- Обучающая выборка (фрагмент)

Тем самым создаются все необходимые и достаточные предпосылки для выявления силы и направления причинно-следственных связей между значениями факторов и результатами их совместного системного воздействия.

1.3. Синтез и верификация статистических и интеллектуальных моделей

Далее необходимо запустить режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей, что показано на рисунке 8.

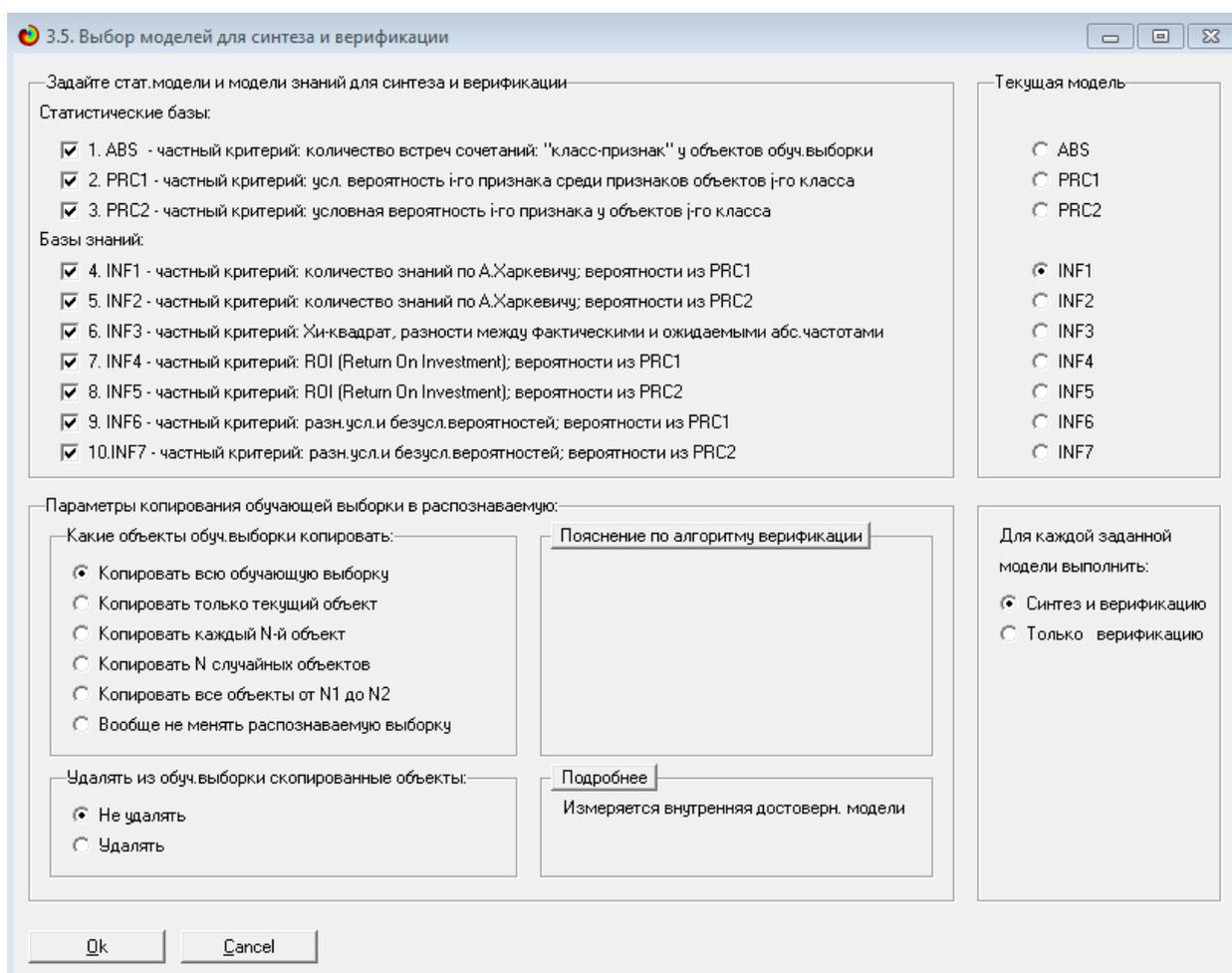


Рисунок 8- Выбор моделей для синтеза и верификации, а также текущей модели

Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 9.

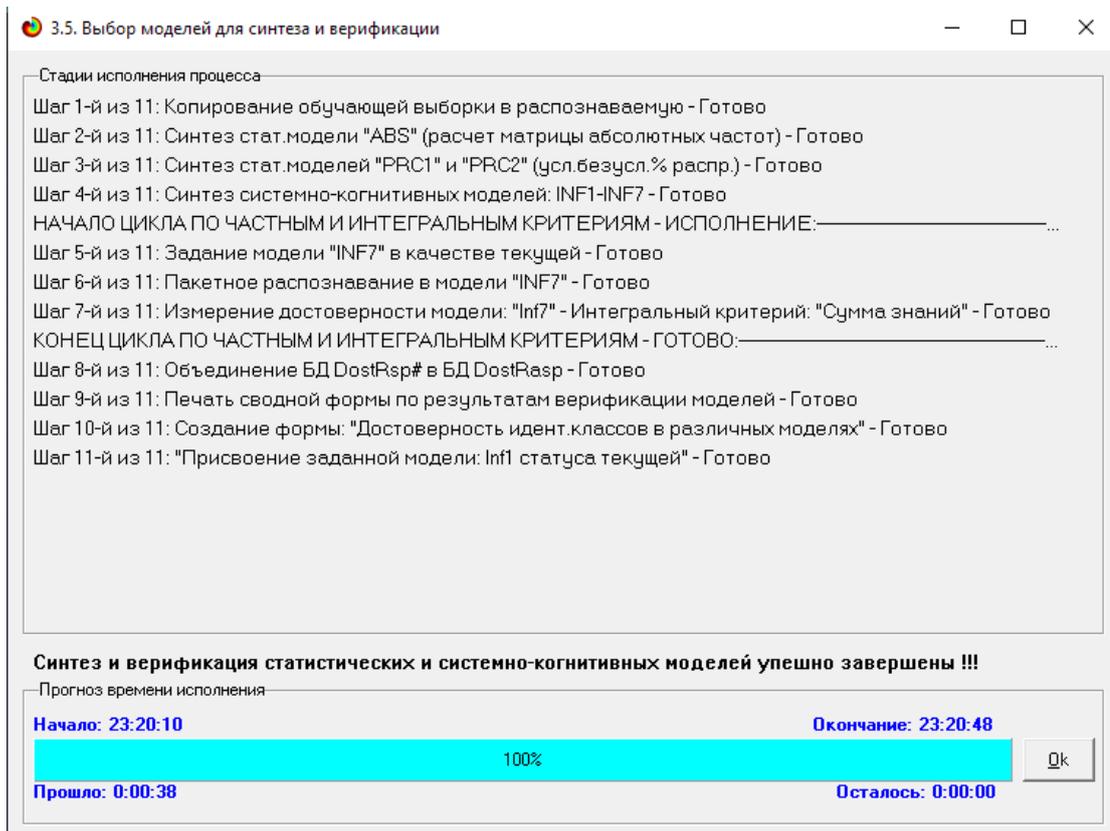


Рисунок 9- Синтез и верификация статистических моделей и моделей знаний

Верификация (оценка достоверности моделей) проводилась на всех примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 8.

1.4. Виды моделей системы «Эйдос»

Рассмотрим решение задачи идентификации на примере модели INF1, в которой рассчитано количество информации по А.Харкевичу, которое мы получаем о принадлежности идентифицируемого объекта к каждому из классов, если знаем, что у этого объекта есть некоторый признак.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот в матрицы условных и безусловных процентных распределений, и матрицы знаний.

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. ТЕКУЩИЙ СТАТУС БАНКРОТ	2. ТЕКУЩИЙ СТАТУС НЕ БАНКРОТ	Сумма	Среднее	Средн. квадр. откл.
1	ДОМА СДАЮТСЯ В СРОК?да		20	20	10.00	14.14
2	ДОМА СДАЮТСЯ В СРОК?нет	24	4	28	14.00	14.14
3	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-1/3-(0.0000000, 2.5000000)	23	23	46	23.00	
4	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-2/3-(2.5000000, 4.5000000)					
5	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-3/3-(4.5000000, 16.0000000)	1	1	2	1.00	
6	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-1/3-(0.0000000, 0.9000000)	23	18	41	20.50	3.54
7	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-2/3-(0.9000000, 1.9000000)		3	3	1.50	2.12
8	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-3/3-(1.9000000, 3.0000000)	1	3	4	2.00	1.41
9	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ?да	24	2	26	13.00	15.56
10	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ?нет		22	22	11.00	15.56
11	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-1/3-(-0.7895236, 0.34...	22	24	46	23.00	1.41
12	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-2/3-(0.3492430, 2.488...	1		1	0.50	0.71
13	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-3/3-(2.4880096, 8.626...	1		1	0.50	0.71
	Сумма числа признаков	120	120	240		
	Среднее	9	9		9.23	
	Среднеквадратичное отклонение	12	10			10.65
	Сумма числа объектов обуч.выборки	24	24	48		

Рисунок 10- Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. ТЕКУЩИЙ СТАТУС БАНКРОТ	2. ТЕКУЩИЙ СТАТУС НЕ БАНКРОТ	Сумма	Среднее	Средн. квадр. откл.
1	ДОМА СДАЮТСЯ В СРОК?да		0.126	0.126	0.063	0.089
2	ДОМА СДАЮТСЯ В СРОК?нет	0.098	-0.229	-0.130	-0.065	0.231
3	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-1/3-(0.0000000, 2.5000000)					
4	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-2/3-(2.5000000, 4.5000000)					
5	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-3/3-(4.5000000, 16.0000000)					
6	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-1/3-(0.0000000, 0.9000000)	0.021	-0.024	-0.003	-0.001	0.032
7	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-2/3-(0.9000000, 1.9000000)		0.126	0.126	0.063	0.089
8	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-3/3-(1.9000000, 3.0000000)	-0.126	0.074	-0.052	-0.026	0.142
9	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ?да	0.112	-0.342	-0.230	-0.115	0.321
10	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ?нет		0.126	0.126	0.063	0.089
11	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-1/3-(-0.7895236, 0.34...	-0.008	0.008	0.000	0.000	0.011
12	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-2/3-(0.3492430, 2.488...	0.126		0.126	0.063	0.089
13	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-3/3-(2.4880096, 8.626...	0.126		0.126	0.063	0.089
	Сумма	0.350	-0.133	0.217		
	Среднее	0.027	-0.010		0.008	
	Среднеквадратичное отклонение	0.071	0.136			0.106

Рисунок 11- Матрица информативности (модель INF1) в битах

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"

Код признака	Наименование описательной шкалы и градации	1. ТЕКУЩИЙ СТАТУС БАНКРОТ	2. ТЕКУЩИЙ СТАТУС НЕ БАНКРОТ	Сумма	Среднее	Средн. квадр. откл.
1	ДОМА СДАЮТСЯ В СРОК? да	-10.000	10.000			14.142
2	ДОМА СДАЮТСЯ В СРОК? нет	10.000	-10.000			14.142
3	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-1/3-(0.0000000, 2.5000000)					
4	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-2/3-(2.5000000, 4.5000000)					
5	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-3/3-(4.5000000, 16.0000000)					
6	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-1/3-(0.0000000, 0.9000000)	2.500	-2.500			3.536
7	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-2/3-(0.9000000, 1.9000000)	-1.500	1.500			2.121
8	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-3/3-(1.9000000, 3.0000000)	-1.000	1.000			1.414
9	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ? да	11.000	-11.000			15.556
10	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ? нет	-11.000	11.000			15.556
11	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-1/3-(0.7895236, 0.34...	-1.000	1.000			1.414
12	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-2/3-(0.3492430, 2.488...	0.500	-0.500			0.707
13	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-3/3-(2.4880096, 8.626...	0.500	-0.500			0.707
	Сумма					
	Среднее					
	Среднеквадратичное отклонение	6.144	6.144			6.020

Рисунок 12- Матрица знаний (модель INF3)

5.5. Модель: "7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. ТЕКУЩИЙ СТАТУС БАНКРОТ	2. ТЕКУЩИЙ СТАТУС НЕ БАНКРОТ	Сумма	Среднее	Средн. квадр. откл.
1	ДОМА СДАЮТСЯ В СРОК? да		1.000	1.000	0.500	0.707
2	ДОМА СДАЮТСЯ В СРОК? нет	0.714	-0.714			1.010
3	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-1/3-(0.0000000, 2.5000000)					
4	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-2/3-(2.5000000, 4.5000000)					
5	КОЛИЧЕСТВО СТРОЯЩИХСЯ ДОМОВ-3/3-(4.5000000, 16.0000000)					
6	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-1/3-(0.0000000, 0.9000000)	0.122	-0.122			0.172
7	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-2/3-(0.9000000, 1.9000000)		1.000	1.000	0.500	0.707
8	КОЛИЧЕСТВО СДАННЫХ ДОМОВ-3/3-(1.9000000, 3.0000000)	-0.500	0.500			0.707
9	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ? да	0.846	-0.846			1.197
10	ДОМА ПРИЗНАНЫ ПРОБЛЕМНЫМИ? нет		1.000	1.000	0.500	0.707
11	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-1/3-(0.7895236, 0.34...	-0.043	0.043			0.061
12	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-2/3-(0.3492430, 2.488...	1.000		1.000	0.500	0.707
13	СООТНОШЕНИЕ СОБСТВЕННОГО И ЗАЕМНОГО КАПИТАЛОВ-3/3-(2.4880096, 8.626...	1.000		1.000	0.500	0.707
	Сумма	3.139	1.861	5.000		
	Среднее	0.241	0.143		0.192	
	Среднеквадратичное отклонение	0.477	0.593			0.528

Рисунок 13 -Матрица знаний (модель INF4)

1.5. Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями, приведены на рисунке 14.

Наименование модели и частного критерия	Интегральный критерий	го ческих ектов орки	Число истинно-положительн... решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных... решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергера
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс. частот с обр...	48	48	5	43		0.527	1.000	0.691
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс. частот по признак...	48	48		48		0.500	1.000	0.667
2. FRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн. частот с о...	48	48	5	43		0.527	1.000	0.691
2. FRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн. частот по при...	48	48		48		0.500	1.000	0.667
3. FRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн. частот с о...	48	48	5	43		0.527	1.000	0.691
3. FRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн. частот по при...	48	48		48		0.500	1.000	0.667
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	48	44	46	2	4	0.957	0.917	0.936
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	48	44	30	18	4	0.710	0.917	0.800
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	48	44	46	2	4	0.957	0.917	0.936
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	48	44	30	18	4	0.710	0.917	0.800
6. INF3 - частный критерий: Хиквадрат_разности между факти...	Семантический резонанс зна...	48	45	45	3	3	0.938	0.938	0.938
6. INF3 - частный критерий: Хиквадрат_разности между факти...	Сумма знаний	48	45	45	3	3	0.938	0.938	0.938
7. INF4 - частный критерий: ROI (Return On Investment): вероятно...	Семантический резонанс зна...	48	44	46	2	4	0.957	0.917	0.936
7. INF4 - частный критерий: ROI (Return On Investment): вероятно...	Сумма знаний	48	46	29	19	2	0.708	0.958	0.814
8. INF5 - частный критерий: ROI (Return On Investment): вероятно...	Семантический резонанс зна...	48	44	46	2	4	0.957	0.917	0.936
8. INF5 - частный критерий: ROI (Return On Investment): вероятно...	Сумма знаний	48	46	29	19	2	0.708	0.958	0.814
9. INF6 - частный критерий: разн.усл.и безуслов. вероятностей; вер...	Семантический резонанс зна...	48	45	45	3	3	0.938	0.938	0.938
9. INF6 - частный критерий: разн.усл.и безуслов. вероятностей; вер...	Сумма знаний	48	45	29	19	3	0.703	0.938	0.804
10. INF7 - частный критерий: разн.усл.и безуслов. вероятностей; ве...	Семантический резонанс зна...	48	45	45	3	3	0.938	0.938	0.938
10. INF7 - частный критерий: разн.усл.и безуслов. вероятностей; ве...	Сумма знаний	48	45	29	19	3	0.703	0.938	0.804

Рисунок 14- Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF3 при интегральных критериях «Семантический резонанс знаний» и «Сумма знаний». При этом точность модели составляет 0,938. Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием, описанным на рисунке 15.

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++"

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++".

положительный псевдопрогноз. Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не относящегося к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

отрицательный псевдопрогноз. представим себе, что мы выбираем кубик с 6 гранями, и модель предсказывает, что не выпадет: 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации.

идеальный прогноз. Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

реальный прогноз. На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вместе число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:

$$L1 = \frac{TP + TN - FP - FN}{TP + TN + FP + FN} \quad (\text{нормировка: } \{-1, +1\})$$

$$L2 = \frac{(1 + (TP + TN - FP - FN) / (TP + TN + FP + FN)) / 2}{(1 + (TP + TN - FP - FN) / (TP + TN + FP + FN)) / 2} \quad (\text{нормировка: } \{0, 1\})$$

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

классическая F-мера достоверности моделей ван Ризбергера (колонка выделена ярко-голубым фоном):

$$F\text{-мера} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Precision = TP / (TP+FP) - точность модели;

Recall = TP / (TP+FN) - полнота модели;

L1-мера проф. Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом сумм уровней сходства (колонка выделена ярко-зеленым фоном):

$$L1\text{-мера} = 2 * (\text{SPrecision} * \text{SRecall}) / (\text{SPrecision} + \text{SRecall})$$

SPrecision = STP / (STP+SFP) - точность с учетом сумм уровней сходства;

SRecall = STP / (STP+SFN) - полнота с учетом сумм уровней сходства;

STP - Сумма модулей сходства истинно-положительных решений; SFN - сумма модулей сходства истинно-отрицательных решений;

SFP - Сумма модулей сходства ложно-положительных решений; SFN - сумма модулей сходства ложно-отрицательных решений.

L2-мера проф. Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом средних уровней сходства (колонка выделена желтым фоном):

$$L2\text{-мера} = 2 * (\text{APrecision} * \text{ARecall}) / (\text{APrecision} + \text{ARecall})$$

APrecision = ATP / (ATP+AFP) - точность с учетом средних уровней сходства;

ARecall = ATP / (ATP+AFN) - полнота с учетом средних уровней сходства;

ATP-SFP/STP - Среднее модулей сходства истинно-положительных решений; AFN-SFN/FN - Среднее модулей сходства истинно-отрицательных решений;

AFP-SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN-SFN/FN - Среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей ван Ризбергера в АСК-анализе и системе "Эйдос" // Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). – С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у. н. п.

Рисунок 15- Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи идентификации

В соответствии с технологией АСК-анализа зададим текущей модель INF3 (режим 5.6) и проведем пакетное распознавание в режиме 4.2.1, как показано на рисунке 17.

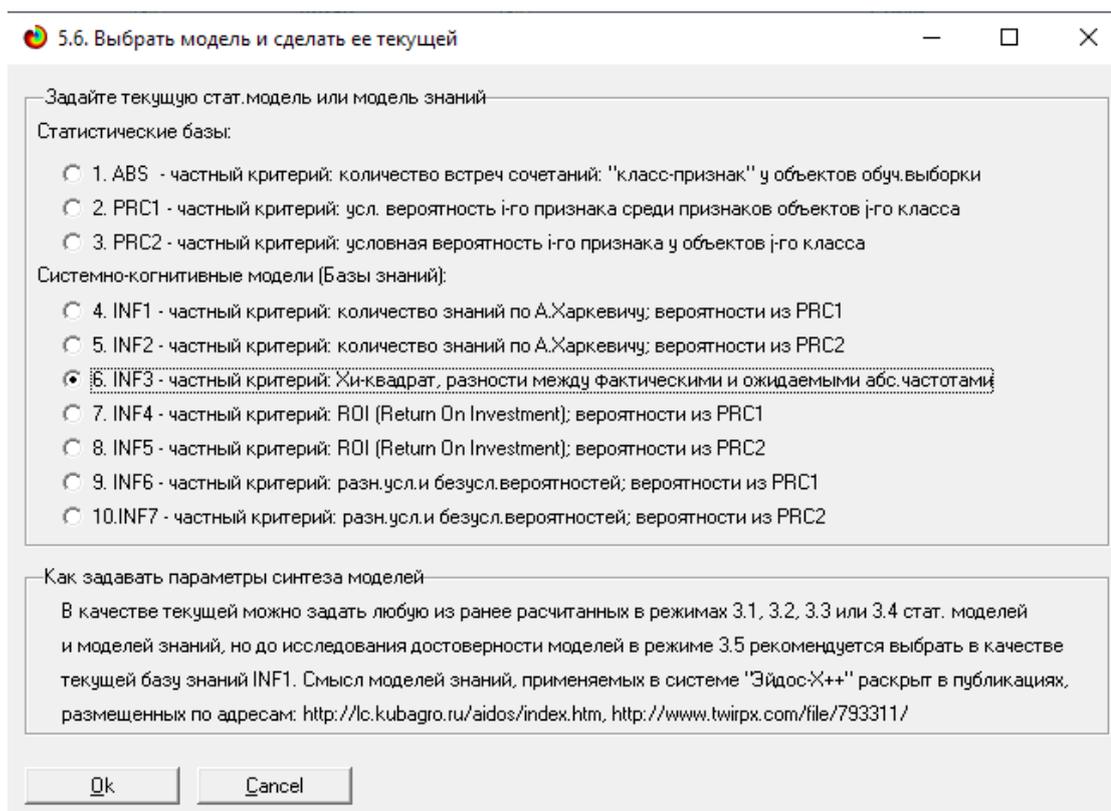


Рисунок 16- Экранная форма режима задания модели в качестве текущей

Следующим шагом необходимо произвести пакетное распознавание в текущей модели. Если распознавание производится впервые, то будет видно сообщение, уведомляющее об этом.

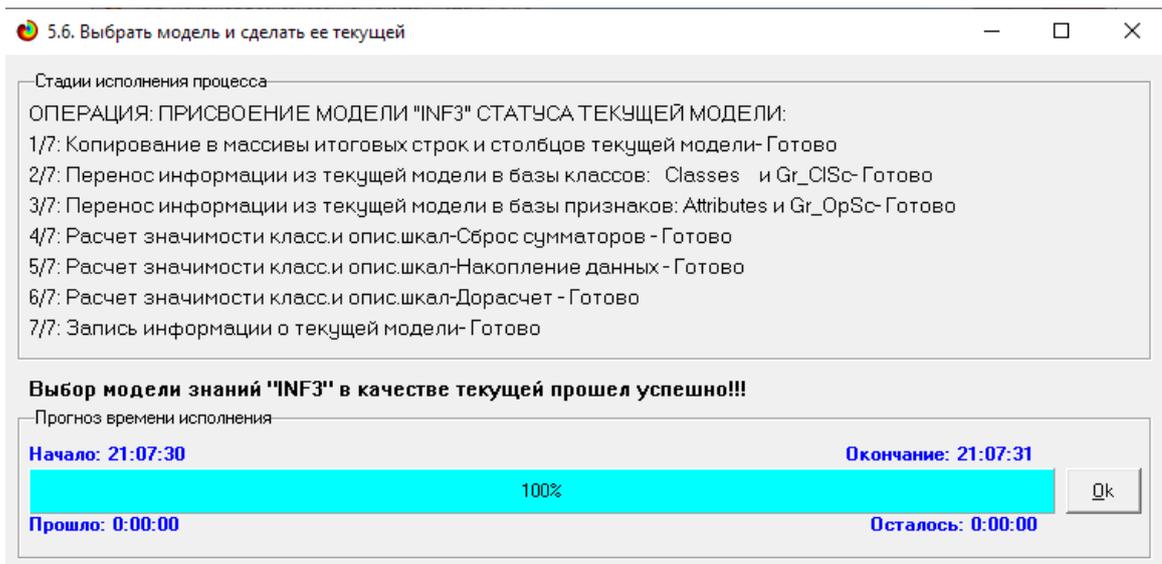


Рисунок 17- Экранная форма режима пакетного распознавания в текущей модели

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных формах, отражающих результаты решения задачи идентификации и прогнозирования.

Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

1. Подробно наглядно: «Объект – классы».
2. Подробно наглядно: «Класс – объекты».
3. Итоги наглядно: «Объект – классы».
4. Итоги наглядно: «Класс – объекты».
5. Подробно сжато: «Объект – классы».
6. Обобщенная форма по достоверности моделей при разных интегральных критериях.
7. Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям.
8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям.
9. Распознавание уровня сходства при разных моделях и интегральных критериях.

10. Достоверность идентификации классов при разных моделях и интегральных критериях.

11. Распределение уровня сходства при разных моделях в интегральных критериях.

12. Объединение в одной БД строк по самым достоверным моделям.

Для наглядного примера, кратко рассмотрим некоторые из них.

На рисунках 18 и 19 приведены примеры прогнозов высокой и низкой достоверности статусов застройщиков в наиболее достоверной модели INF3 на основе описательных шкал.

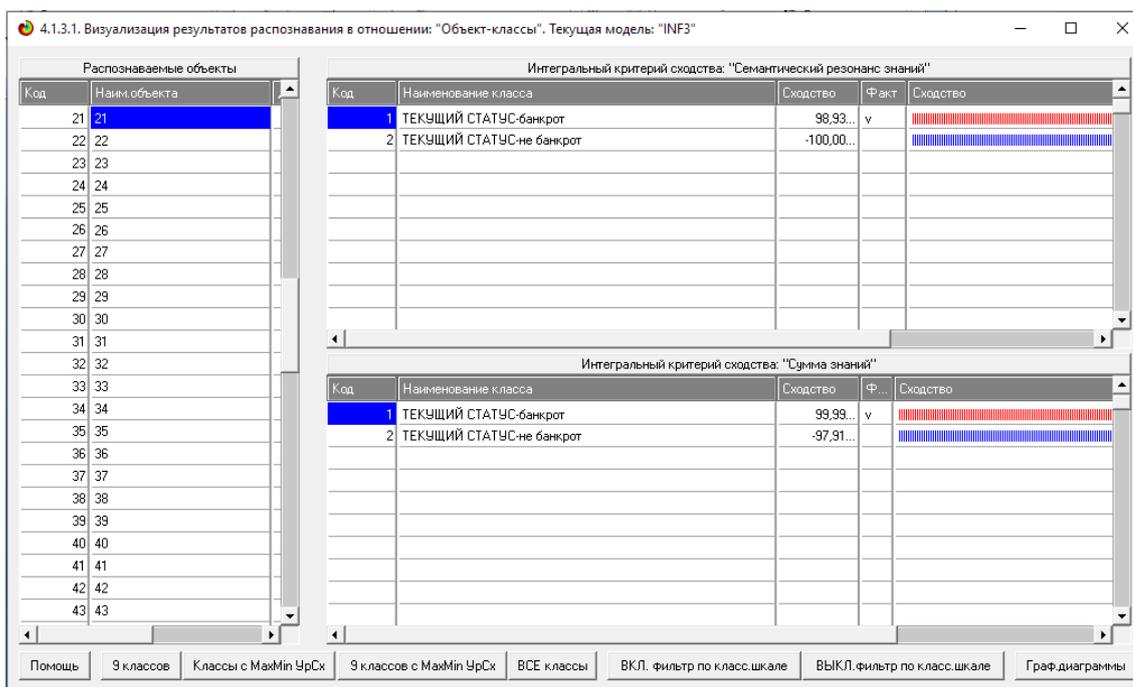


Рисунок 18- Пример идентификации классов в модели INF3

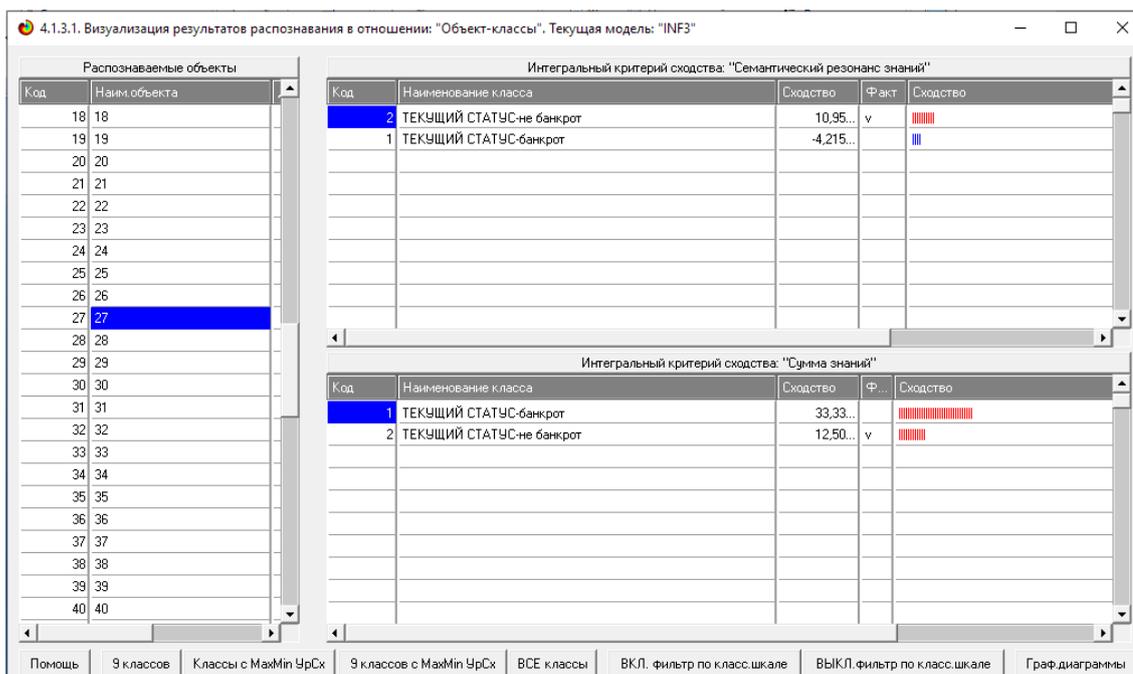


Рисунок 19- Пример идентификации классов в модели INF3

2.2. Когнитивные функции

Далее стоит рассмотреть режим работы 4.5, в котором реализована возможность визуализации когнитивных функций для любых моделей и любых сочетаний классификационных и описательных шкал. Начало работы в этом режиме начинается с понятия когнитивных функций, а также предоставления возможности перехода на статьи по данной проблематике и скачивания публикаций, как это продемонстрировано на рисунке 20.

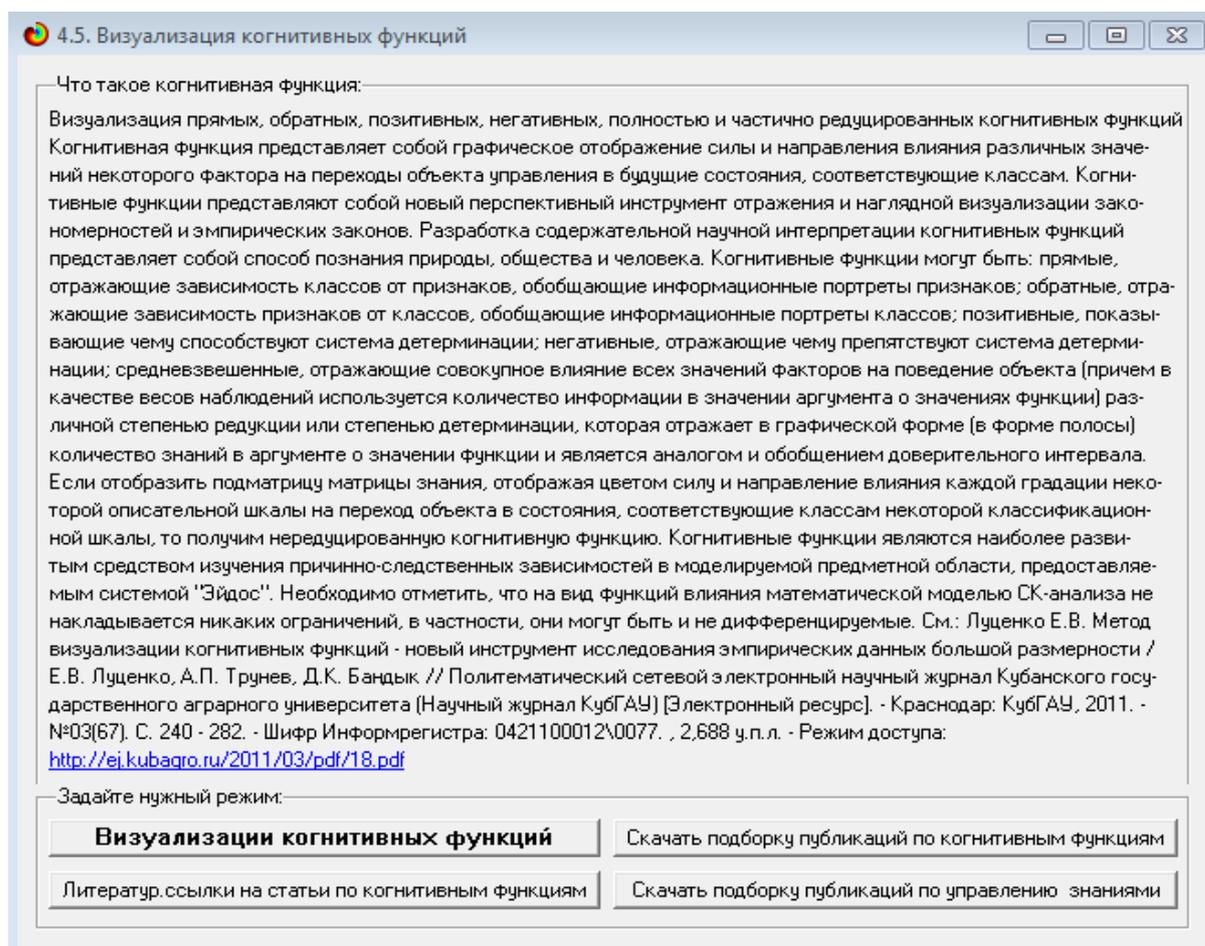


Рисунок 20- Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

Применительно к задаче, рассматриваемой в данной работе, когнитивная функция показывает, какое количество информации содержится в различных значениях факторов о том, что объект моделирования перейдет в те или иные будущие состояния. На рисунке 21 приведены визуализации когнитивных функций данного приложения для модели INF3.

На рисунке 21 показано, что, если дома сдаются в срок, то строительная организация не является банкротом.

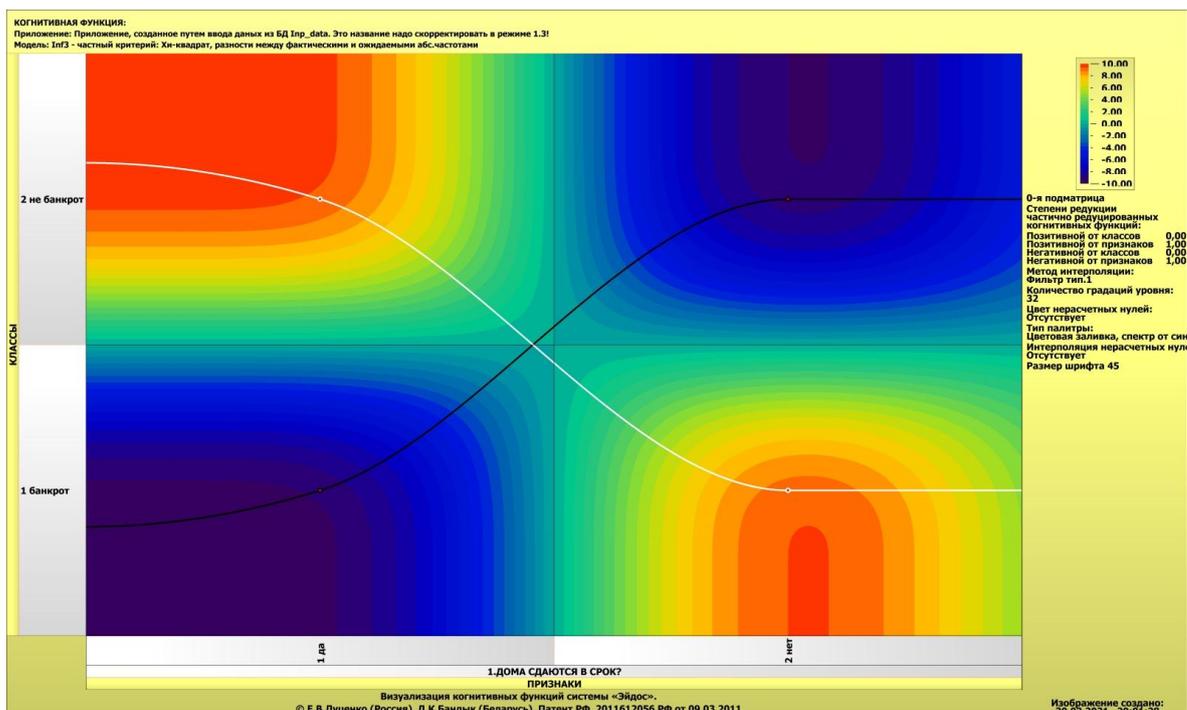


Рисунок 21-Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF3

2.3. SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных.

Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOT-анализа выводились в форме информационных портретов. В версии системы под MS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 22).

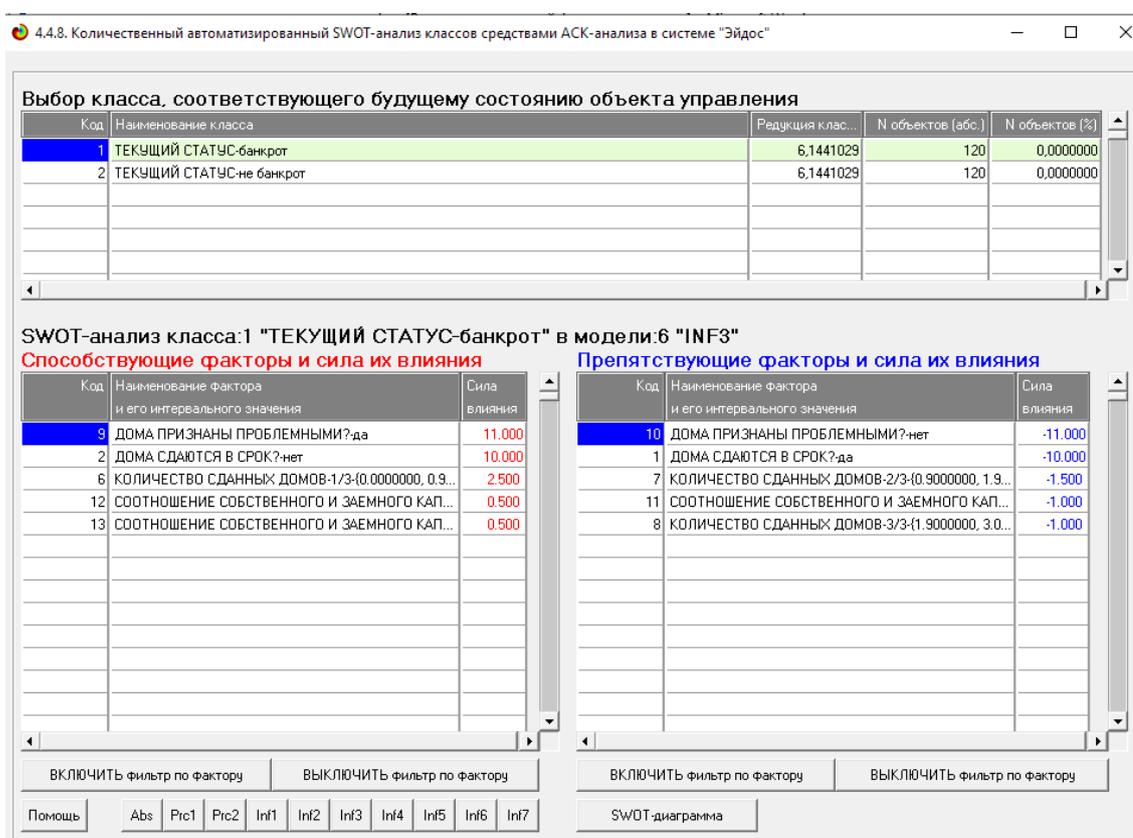


Рисунок 22-Пример SWOT-Матрицы в модели INF3

Следующим шагом является построение SWOT-диаграммы, для наглядной демонстрации работы, продемонстрированной на рисунке 23.

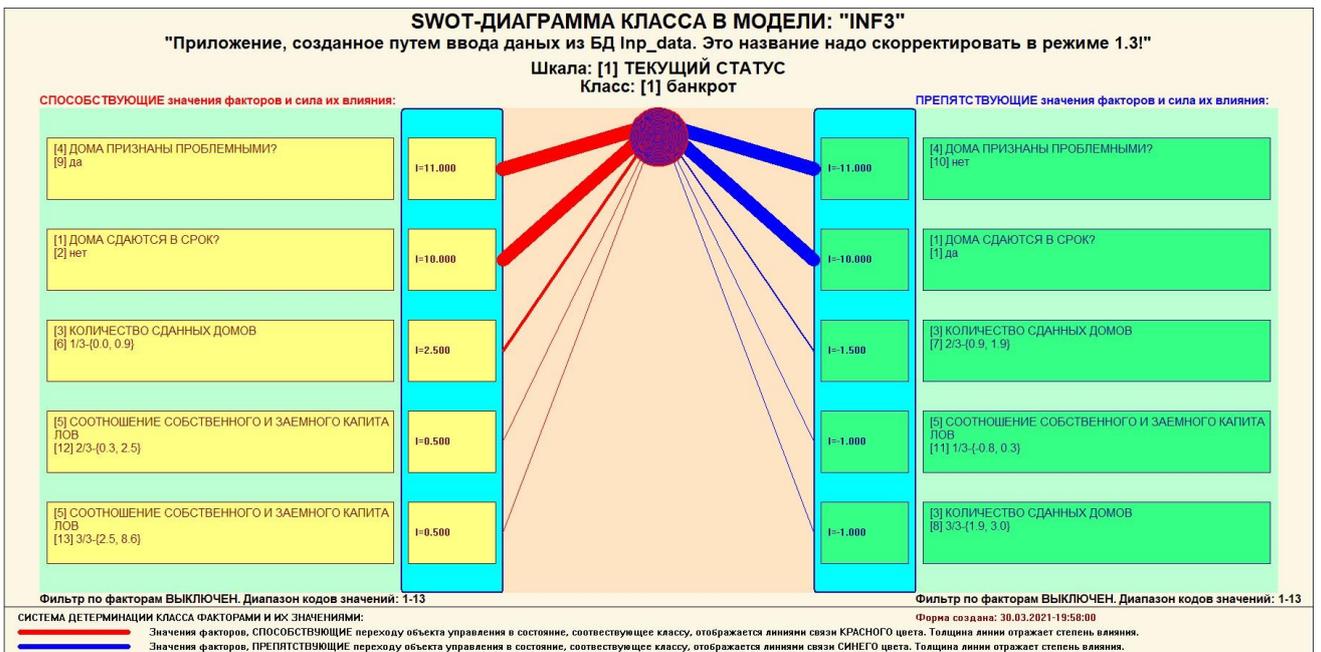


Рисунок 23– SWOT-матрица

На диаграмме показываются способствующие и препятствующие факторы. К наиболее способствующим факторам относятся:

- Присутствие проблемных домов;
- Длительный срок сдачи дома;
- Количество сданных домов [6]1/3-{0.0,0.9}, где $i=2.500$;
- Верхний интервал соотношения собственного и заемного капиталов [12]2/3-{0.3,2.5}, где $i=0.500$;
- Нижний интервал соотношения собственного и заемного капиталов [13]3/3-{2.5,8.6}, где $i=0.500$;

К наиболее препятствующим относятся:

- Отсутствие значений проблемных домов;
- Своевременная сдача дома в эксплуатацию;
- Количество сданных домов [7]2/3-{0.9,1.9}, где $i=1.500$;
- Соотношение собственного и заемного капиталов [11]1/3-{0.8,0.3}, где $i=1.000$;
- Количество сданных домов [8]3/3-{1.9,3.0}, где $i=1.000$.

2.4. Нелокальные нейронные сети

Каждому классу системно-когнитивной модели соответствует нелокальный нейрон, совокупность которых образует не локальную нейронную сеть. Рассмотрим пример на рисунке 24.

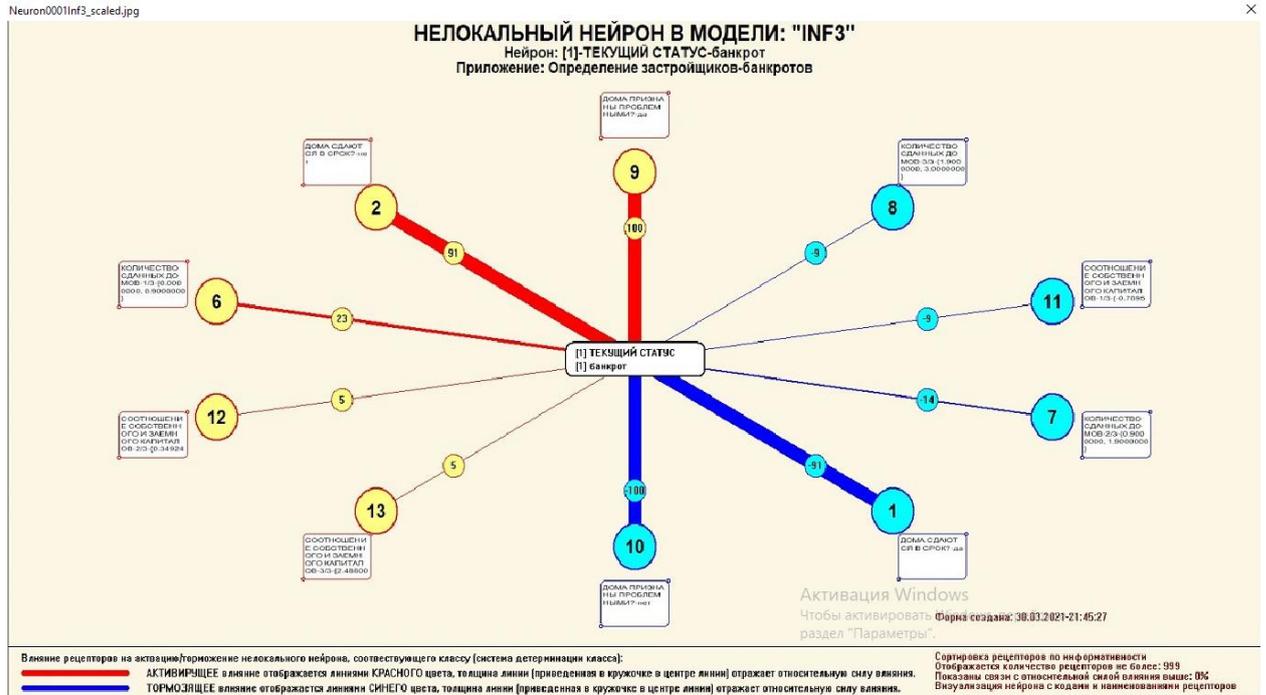


Рисунок 24- Нейрон в модели INF3

Благодаря данному нейрону видно, как различные факторы влияют на статус застройщика-банкрота, какие оказывают положительное влияние, а какие отрицательное.

2.5. Паретто-подмножества нелокальной нейронной сети

Если модель объекта достаточно адекватна, то ее исследование корректно считать исследованием самого моделируемого объекта. В нашем случае это именно так.

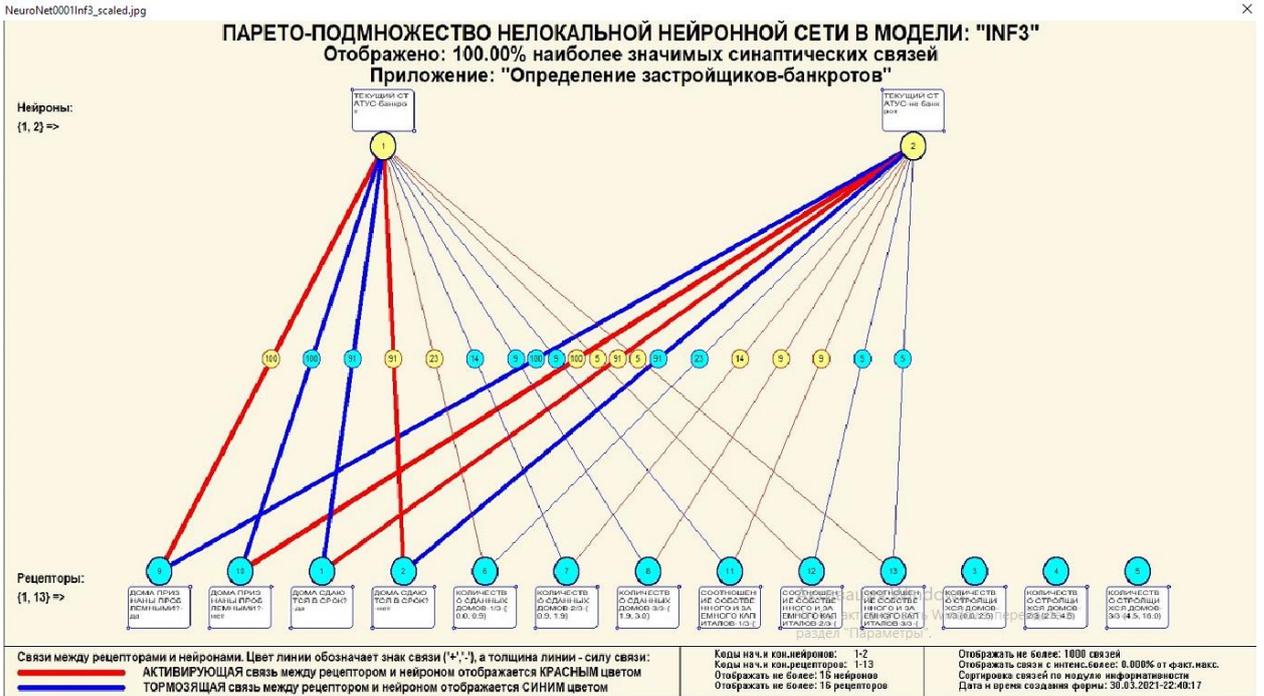


Рисунок 25- Паретто - подмножество нелокальной нейронной сети в модели INF3

На всех формах красным показаны значения факторов, способствующие достижению этого результата, а синим – препятствующие. Сила влияния (значимость) выражена в битах и отображается толщиной линии. Приведем Паретто-кривую значимости градаций описательных шкал, т.е. значений экологических факторов (рисунок 26).

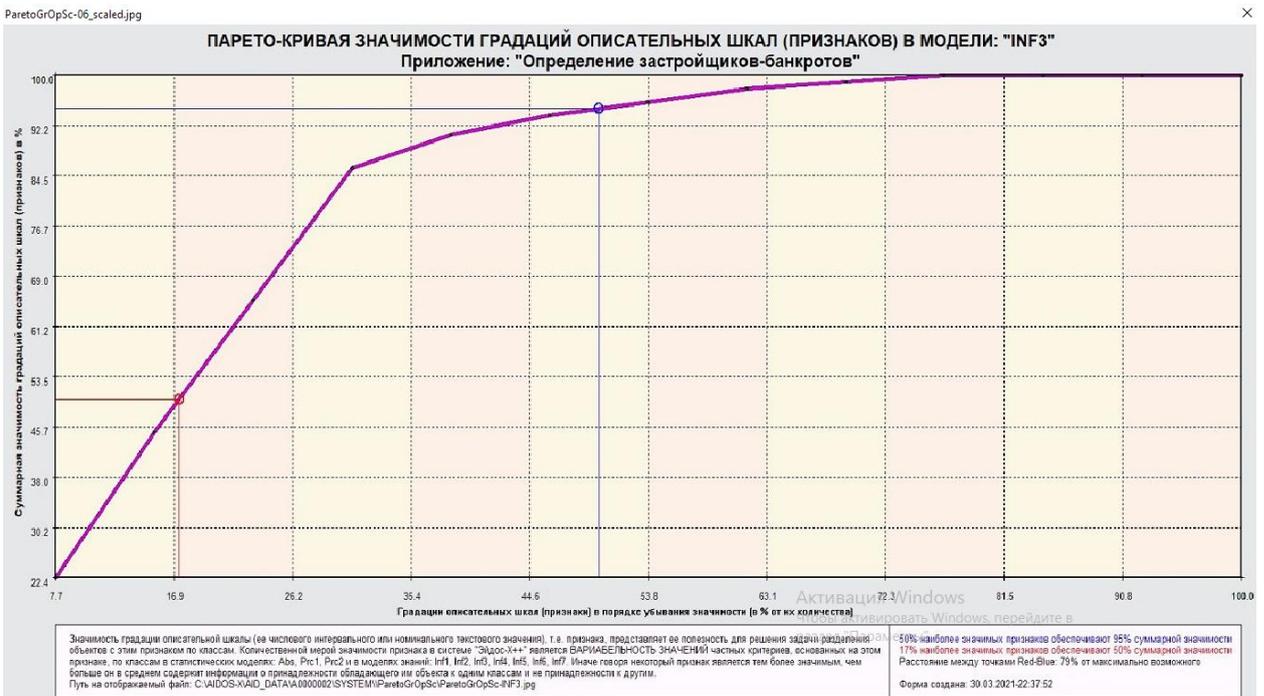


Рисунок 26 – Паретто-кривая значимости значений факторов

Из нее хорошо видно, что в модели Inf3 около 50 процентов суммарного влияния всех значений факторов обусловлено всего 17 процентами от их числа, а остальные 83 процентов все вместе играют примерно такую же роль, как этих 17 наиболее сильно влияющих. По-видимому, на этих наиболее сильно влияющих значениях факторов и следует сосредоточить основное внимание при планировании и проведении экологических мероприятий.

2.6. Когнитивные диаграммы

Когнитивная диаграмма классов отражает их сходство и различие по системе обуславливающих их значений факторов.

Необходимо отметить, что величина сходства и различия в когнитивных диаграммах получена в результате расчета матриц сходства на основе моделей знаний, а не в результате не формализуемых экспертных оценок на основе интуиции и профессиональной компетенции.

На рисунке 27 приведен фрагмент интегральной когнитивной карты, представляющей собой 3D визуализацию фрагмента нелокальной нейронной сети, нейроны которой соединены когнитивной диаграммой, как и значения факторов.

Ниже приведена когнитивная диаграмма классов, на рисунке 28, содержательно отражающая сходство-различие двух SWOT-диаграмм классов. После чего на рисунке 29 приведена когнитивная диаграмма значений факторов, содержательно отражающая сходство-различие двух SWOT-диаграмм значений факторов.

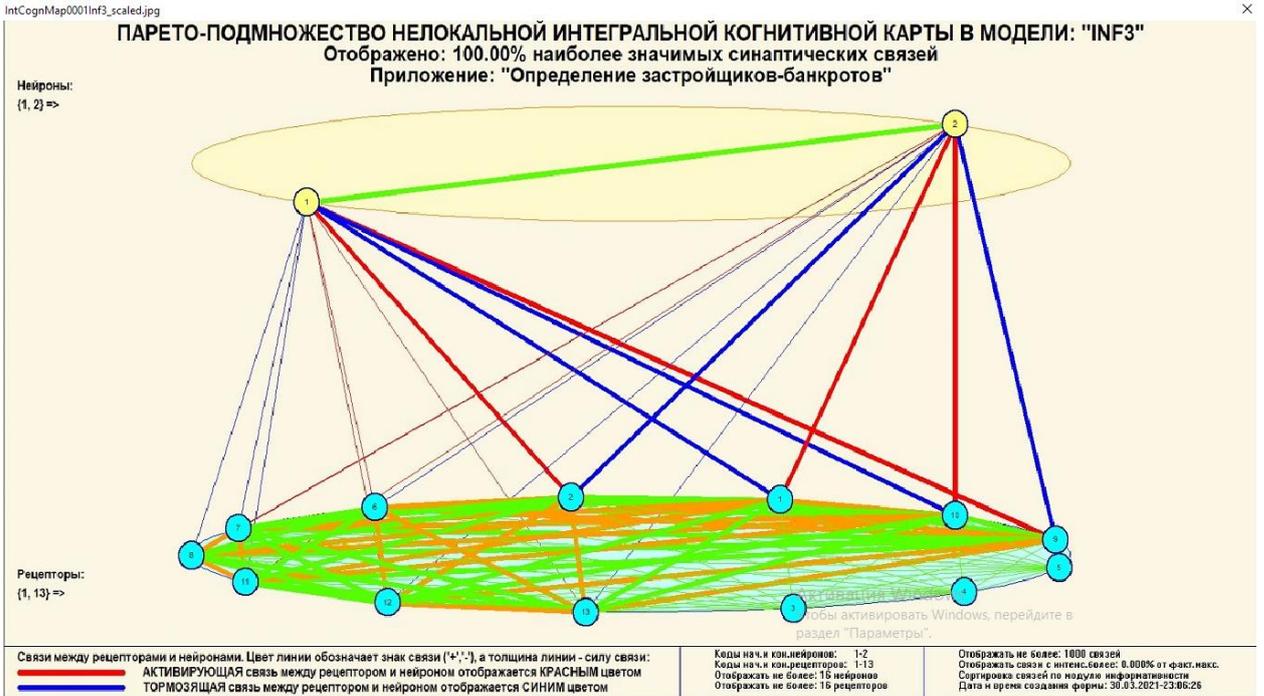


Рисунок 27 – Дендрограмма когнитивной кластеризации значений факторов и график изменения межкластерных расстояний

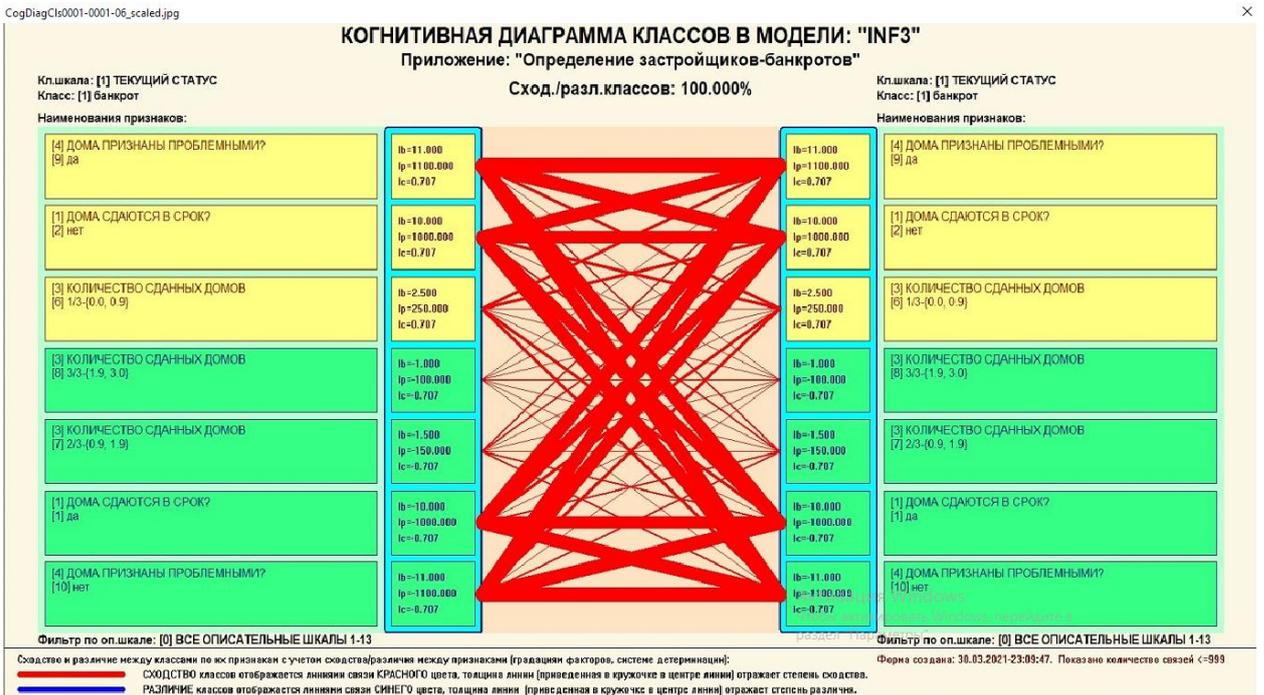


Рисунок 28 – Когнитивная диаграмма классов



Рисунок 29 – Когнитивная диаграмма значений факторов

ЗАКЛЮЧЕНИЕ

В связи с существованием множества систем искусственного интеллекта возникает необходимость сопоставимой оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных, для чего была использована информация ресурсов единой информационной системы жилищного строительства, единого ресурса застройщиков, каталога организаций.

В данной работе приводится развернутый пример использования базы данных для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системы искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказалась модель INF3, основанная на семантической мере целесообразности информации А.Харкевича при интегральных критериях «Семантический резонанс знаний » и «Сумма знаний». При этом точность модели составляет 0,938, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием.

На основе исходных данных, рассмотренных в данной работе, построены модели прогнозирования с помощью АСК-анализа и реализующей его системы «Эйдос – X ++».

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.

2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.

3. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.

4. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических,

технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

6. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://ic.kubagro.ru/>, свободный. - Загл. с экрана. Яз. рус.

7. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 48 – 77. – Шифр Информрегистра: 0420900012\0110, IDA [article ID]: 0540910004. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.