

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ  
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет заочного обучения  
Кафедра компьютерных технологий и систем

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

**к курсовой работе**

по дисциплине: Системы представления и приобретения знаний  
на тему:

**Прогнозирование наиболее опасных районов в Бостоне на основе АСК-  
анализа преступлений**

выполнил студент группы ИТ31841

**Дробященко Виталий Алексеевич**

Допущена к защите

Руководитель проекта Луценко Евгений Вениаминович, д.э.н., к.т.н.,  
профессор

(подпись, расшифровка подписи)

Нормоконтролер Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор

(подпись, расшифровка подписи)

Защищена \_\_\_\_\_

Оценка \_\_\_\_\_

\_\_\_\_\_  
(дата)

Члены комиссии \_\_\_\_\_ В.И. Лойко  
\_\_\_\_\_ Е.В. Луценко  
\_\_\_\_\_ В.Н. Лаптев

(подпись, дата, расшифровка подписи)

Краснодар  
2019 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ  
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики  
Кафедра компьютерных технологий и систем

**УТВЕРЖДАЮ:**

**Зав. кафедрой КТС \_\_\_\_\_ В. И. Лойко**

**ЗАДАНИЕ**

**на курсовую работу**

Студенту: ИТ31841 группы 2 курса  
Факультета прикладной информатики  
Специальности: 09.04.02 «Информационные системы и технологии»  
(шифр)

**Дробященко Виталию Алексеевичу**  
(Ф.И.О.)

Тема проекта: **Прогнозирование наиболее опасных районов в Бостоне на основе АСК-анализа преступлений**

Содержание задания: Проанализировать современные методы и средства формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами

Объем работы:

- а) пояснительная записка к работе \_\_\_\_\_ 47 листов формата А4  
б) графическая часть \_\_\_\_\_ лист формата А4

Рекомендуемая литература: Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос- X++» 2017 г.

Срок выполнения проекта: с “ \_\_\_\_\_ ” \_\_\_\_\_ по “ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ г.

Срок защиты: “ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ г.

Дата выдачи задания: “ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ г.

Дата сдачи проекта на кафедру: “ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ г.

Руководитель проекта: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор

(подпись, Ф.И.О., звание, степень)

Задание принял студент \_\_\_\_\_

(подпись, дата)

## Реферат

Пояснительная записка курсового проекта (работы) 39 с., 28 рис., 2 табл., 12 источников.

Ключевые слова: БОСТОН, РАЙОН ПРЕСТУПЛЕНИЙ, ВРЕМЯ ПРЕСТУПЛЕНИЙ, РАЙОН, ПРЕСТУПЛЕНИЯ В БОСТОНЕ, ОПАСНЫЙ РАЙОН, ПРОГНОЗ, ЭЙДОС X, СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, ГРАДАЦИИ, КЛАССИФИКАЦИОННЫЕ ШКАЛЫ, ОПИСАТЕЛЬНЫЕ ШКАЛЫ

В курсовом проекте рассматриваются вопросы прогнозирования наиболее опасных районов преступлений в Бостоне путем применения АСК-анализа и системы Эйдос на основе данных репозитория KAGGLE.

## Оглавление

1. Синтез и верификация моделей.....	6
1.1. Описание решения .....	6
1.2. Подготовка и нормализация исходных данных.....	6
1.3. Синтез и верификация статистических и интеллектуальных моделей.....	14
1.4. Верификация статистических и системно-когнитивных моделей....	18
2. Решение задач идентификации в наиболее достоверной модели.....	24
2.1. Выбор наиболее достоверной модели и присвоение ей статуса текущей .....	24
2.2. Прогнозирование (диагностика, классификация, распознавание, идентификация) .....	25
2.3. Визуализация когнитивных функций .....	28
2.4. Нелокальные нейроны и нейронная сеть.....	31
2.5. SWOT и PEST матрицы и диаграммы.....	35
Список литературы .....	38

## **Введение**

Создание систем искусственного интеллекта является одним из важных и перспективных направлений развития современных информационных технологий. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи по определению наиболее опасного времени и места преступления в Бостоне.

Для достижения поставленной цели необходимы свободный доступ к тестовым исходным данным и методика, которая поможет преобразовать эти данные в форму, которая необходима для работы в системе искусственного интеллекта. Удачным выбором является сборник баз данных Kaggle.

В данной курсовой работе использована база данных «crime boston» из банка исходных данных по задачам искусственного интеллекта – репозитория Kaggle.

Для решения задачи используем стандартные возможности Microsoft Office Word и 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 из csv-формата в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы "Aidos".
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

### **1.2. Подготовка и нормализация исходных данных**

С электронного ресурса <https://www.kaggle.com/AnalyzeBoston/crimes-in-boston> возьмем базу данных: crime.csv

В файле – объединение срезов базы данных преступлений в определенные моменты времени. Он содержит:

- INCIDENT\_NUMBER
- OFFENSE\_CODE
- OFFENSE\_CODE\_GROUP
- OFFENSE\_DESCRIPTION
- DISTRICT
- REPORTING\_AREA
- SHOOTING
- OCCURRED\_ON\_DATE
- YEAR
- MONTH
- DAY\_OF\_WEEK
- HOUR
- UCR\_PART

- STREET
- Lat
- Long
- Location

INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE
1182070945,00619,	Larceny,LARCENY ALL OTHERS,D14,808,,		2018-09-02 13:00:00,2018,9,Sunday,13,Part One,LINCOLN ST,42.35779134,-71.13937053,"(4				
1182070943,01402,	Vandalism,VANDALISM,C11,347,,		2018-08-21 00:00:00,2018,8,Tuesday,0,Part Two,HECLA ST,42.30682138,-71.06030035,"(42.306821				
1182070941,03410,	Towed,TOWED MOTOR VEHICLE,D4,151,,		2018-09-03 19:27:00,2018,9,Monday,19,Part Three,CAZENOVE ST,42.34658879,-71.072429				
1182070940,03114,	Investigate Property,INVESTIGATE PROPERTY,D4,272,,		2018-09-03 21:16:00,2018,9,Monday,21,Part Three,NEWCOMB ST,42.334181				
1182070938,03114,	Investigate Property,INVESTIGATE PROPERTY,B3,421,,		2018-09-03 21:05:00,2018,9,Monday,21,Part Three,DELHI ST,42.27536542,-71.05860832				
1182070936,03820,	Motor Vehicle Accident Response,M/V ACCIDENT INVOLVING PEDESTRIAN - INJURY,C11,398,,		2018-09-03 21:09:00,2018,9,Monday				
1182070933,00724,	Auto Theft,AUTO THEFT,B2,330,,		2018-09-03 21:25:00,2018,9,Monday,21,Part One,NORMANDY ST,42.30607218,-71.08273260,"(42.3				
1182070932,03301,	Verbal Disputes,VERBAL DISPUTE,B2,584,,		2018-09-03 20:39:37,2018,9,Monday,20,Part Three,LAWN ST,42.32701648,-71.10555088,"				
1182070931,00301,	Robbery,ROBBERY - STREET,C6,177,,		2018-09-03 20:48:00,2018,9,Monday,20,Part One,MASSACHUSETTS AVE,42.33152148,-71.07085				
1182070929,03301,	Verbal Disputes,VERBAL DISPUTE,C11,364,,		2018-09-03 20:38:00,2018,9,Monday,20,Part Three,LESLIE ST,42.29514664,-71.05860832				
1182070928,03301,	Verbal Disputes,VERBAL DISPUTE,C6,913,,		2018-09-03 19:55:00,2018,9,Monday,19,Part Three,OCEAN VIEW DR,42.31957856,-71.040				
1182070927,03114,	Investigate Property,INVESTIGATE PROPERTY,C6,936,,		2018-09-03 20:19:00,2018,9,Monday,20,Part Three,DALESSIO CT,42.3401146				
1182070923,03108,	Fire Related Reports,"FIRE REPORT - HOUSE, BUILDING, ETC.",D4,139,,		2018-09-03 19:58:00,2018,9,Monday,19,Part Three,MARLBOR				
1182070922,02647,	Other,THREATS TO DO BODILY HARM,B3,429,,		2018-09-03 20:39:00,2018,9,Monday,20,Part Two,WOODROW AVE,42.28647012,-71.0				
1182070921,03201,	Property Lost,PROPERTY - LOST,B3,469,,		2018-09-02 14:00:00,2018,9,Sunday,14,Part Three,MULVEY ST,42.27924052,-71.09667382,"(				
1182070920,03006,	Medical Assistance,SICK/INJURED/MEDICAL - PERSON,,		2018-09-03 19:43:00,2018,9,Monday,19,Part Three,,42.35287456,-71.0738				
1182070919,03301,	Verbal Disputes,VERBAL DISPUTE,C11,341,,		2018-09-03 18:52:00,2018,9,Monday,18,Part Three,STONEHURST ST,42.30526428,-71.06				
1182070918,03305,	Assembly or Gathering Violations,DEMONSTRATIONS/RIOT,D4,130,,		2018-09-03 17:00:00,2018,9,Monday,17,Part Three,HUNTINGT				
1182070917,02647,	Other,THREATS TO DO BODILY HARM,B2,901,,		2018-09-03 19:52:00,2018,9,Monday,19,Part Two,HORADAN WAY,42.33371742,-71.09				
1182070915,00614,	Larceny From Motor Vehicle,LARCENY THEFT FROM MV - NON-ACCESSORY,B2,181,,		2018-09-02 18:00:00,2018,9,Sunday,18,Part On				
1182070913,03006,	Medical Assistance,SICK/INJURED/MEDICAL - PERSON,,		2018-09-03 18:46:00,2018,9,Monday,18,Part Three,WOLCOTT,-1.0000000				
1182070911,03801,	Motor Vehicle Accident Response,M/V ACCIDENT - OTHER,A1,69,,		2018-09-03 18:30:00,2018,9,Monday,18,Part Three,BEACON ST,4				
1182070910,03006,	Medical Assistance,SICK/INJURED/MEDICAL - PERSON,B3,434,,		2018-09-03 18:42:00,2018,9,Monday,18,Part Three,CAPEN ST,42.28				
1182070909,03803,	Motor Vehicle Accident Response,M/V ACCIDENT - PERSONAL INJURY,E5,550,,		2018-09-03 18:33:00,2018,9,Monday,18,Part Three,				
1182070908,00522,	Residential Burglary,BURGLARY - RESIDENTIAL - NO FORCE,B2,911,,		2018-09-03 18:38:00,2018,9,Monday,18,Part One,ANNUNCIATI				
1182070906,03831,	Motor Vehicle Accident Response,M/V - LEAVING SCENE - PROPERTY DAMAGE,,		2018-09-03 18:20:00,2018,9,Monday,18,Part Thr				
1182070905,03006,	Medical Assistance,SICK/INJURED/MEDICAL - PERSON,D4,172,,		2018-09-03 18:50:00,2018,9,Monday,18,Part Three,MASSACHUSETT				
1182070904,00802,	Simple Assault,ASSAULT SIMPLE - BATTERY,C11,242,,		2018-09-03 18:34:00,2018,9,Monday,18,Part Two,ANNAPOLIS ST,42.31731905,				
1182070904,02007,	Restraining Order Violations,VIOL. OF RESTRAINING ORDER W NO ARREST,C11,242,,		2018-09-03 18:34:00,2018,9,Monday,18,Part T				
1182070903,02900,	Other,VAL - VIOLATION OF AUTO LAW - OTHER,B3,463,,		2018-09-03 18:55:00,2018,9,Monday,18,Part Two,BLUE HILL AVE,42.295903				
1182070901,02907,	Violations,VAL - OPERATING AFTER REV/SUSP.,B3,428,,		2018-09-03 18:41:00,2018,9,Monday,18,Part Two,CLARKWOOD ST,42.28013				
1182070900,02629,	Harassment,HARASSMENT,B3,464,,		2018-09-03 18:17:00,2018,9,Monday,18,Part Two,HANSBOROUGH ST,42.28810362,-71.09153335				
1182070898,00802,	Simple Assault,ASSAULT SIMPLE - BATTERY,C11,351,,		2018-09-03 19:11:00,2018,9,Monday,19,Part Two,SALISBURY PARK,42.2992837				

Рисунок 1 – Фрагмент данных базы crime.csv

В качестве разделителей используются запятые. Из всех вышеперечисленных характеристик нас интересуют:

- ID
- DAY\_OF\_WEEK
- MONTH
- YEAR

- HOUR
- OFFENSE\_DESCRIPTION
- STREET

Это нам позволит понять, в каких районах наиболее опасно и в какое время, а также поможет полицейским направлять силы на патрулирования, для избегания данных преступлений.

С исходными данными были выполнены следующие преобразования:

- файлы с данными были вырезаны для получения наиболее свежих данных в выборке, объем выборки: 1500 строк;
- замена всех запятых на точку с запятой;
- замена точек на запяты;
- удалено поле date и были добавлены два новые поля: день недели и час
- добавлено поле id с инкрементальной нумерацией;
- столбцы были поменяны местами, для наилучшего их восприятия;
- файл сохранен с именем inp\_data.xlsx.

Фрагмент подготовленной и нормализованной выборке приведен на рисунке (Рисунок 2).

	A	B	C	D	E	F
1	ID	DAY_OF_WEEK	MONTH	HOUR	OFFENSE_DESCRIPTION	STREET
2	1	Sunday	9	13	LARCENY ALL OTHERS	LINCOLN ST
3	2	Tuesday	8	0	VANDALISM	HECLA ST
4	3	Monday	9	19	TOWED MOTOR VEHICLE	CAZENOVE ST
5	4	Monday	9	21	INVESTIGATE PROPERTY	NEWCOMB ST
6	5	Monday	9	21	INVESTIGATE PROPERTY	DELHI ST
7	6	Monday	9	21	M/V ACCIDENT INVOLVING PEDESTRIAN - INJURY	TALBOT AVE
8	7	Monday	9	21	AUTO THEFT	NORMANDY ST
9	8	Monday	9	20	VERBAL DISPUTE	LAWN ST
10	9	Monday	9	20	ROBBERY - STREET	MASSACHUSETTS AVE
11	10	Monday	9	20	VERBAL DISPUTE	LESLIE ST
12	11	Monday	9	19	VERBAL DISPUTE	OCEAN VIEW DR
13	12	Monday	9	20	INVESTIGATE PROPERTY	DALESSIO CT
14	13	Monday	9	19	FIRE REPORT - HOUSE, BUILDING, ETC.	MARLBOROUGH ST
15	14	Monday	9	20	THREATS TO DO BODILY HARM	WOODROW AVE
16	15	Sunday	9	14	PROPERTY - LOST	MULVEY ST
17	16	Monday	9	19	SICK/INJURED/MEDICAL - PERSON	
18	17	Monday	9	18	VERBAL DISPUTE	STONEHURST ST
19	18	Monday	9	17	DEMONSTRATIONS/RIOT	HUNTINGTON AVE
20	19	Monday	9	19	THREATS TO DO BODILY HARM	HORADAN WAY
21	20	Sunday	9	18	LARCENY THEFT FROM MV - NON-ACCESSORY	SHIRLEY ST
22	21	Monday	9	18	SICK/INJURED/MEDICAL - PERSON	WOLCOTT
23	22	Monday	9	18	M/V ACCIDENT - OTHER	BEACON ST
24	23	Monday	9	18	SICK/INJURED/MEDICAL - PERSON	CAPEN ST
25	24	Monday	9	18	M/V ACCIDENT - PERSONAL INJURY	WASHINGTON ST
26	25	Monday	9	18	BURGLARY - RESIDENTIAL - NO FORCE	ANNUNCIATION RD
27	26	Monday	9	18	M/V - LEAVING SCENE - PROPERTY DAMAGE	
28	27	Monday	9	18	SICK/INJURED/MEDICAL - PERSON	MASSACHUSETTS AVE
29	28	Monday	9	18	ASSAULT SIMPLE - BATTERY	ANNAPOLIS ST
30	29	Monday	9	18	VIOL. OF RESTRAINING ORDER W NO ARREST	ANNAPOLIS ST
31	30	Monday	9	18	VAL - VIOLATION OF AUTO LAW - OTHER	BLUE HILL AVE
32	31	Monday	9	18	VAL - OPERATING AFTER REV/SUSP.	CLARKWOOD ST
33	32	Monday	9	18	HARASSMENT	HANSBOROUGH ST
34	33	Monday	9	19	ASSAULT SIMPLE - BATTERY	SALISBURY PARK

Рисунок 1 – Фрагмент данных обучающей выборки

Полученный файл необходимо разместить в папке «Эйдос»: \idos-X\AID\_DATA\Inp\_data\. Автоматизированная формализация предметной области происходит путем импорта исходных данных из внешних баз данных в систему «Эйдос».

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

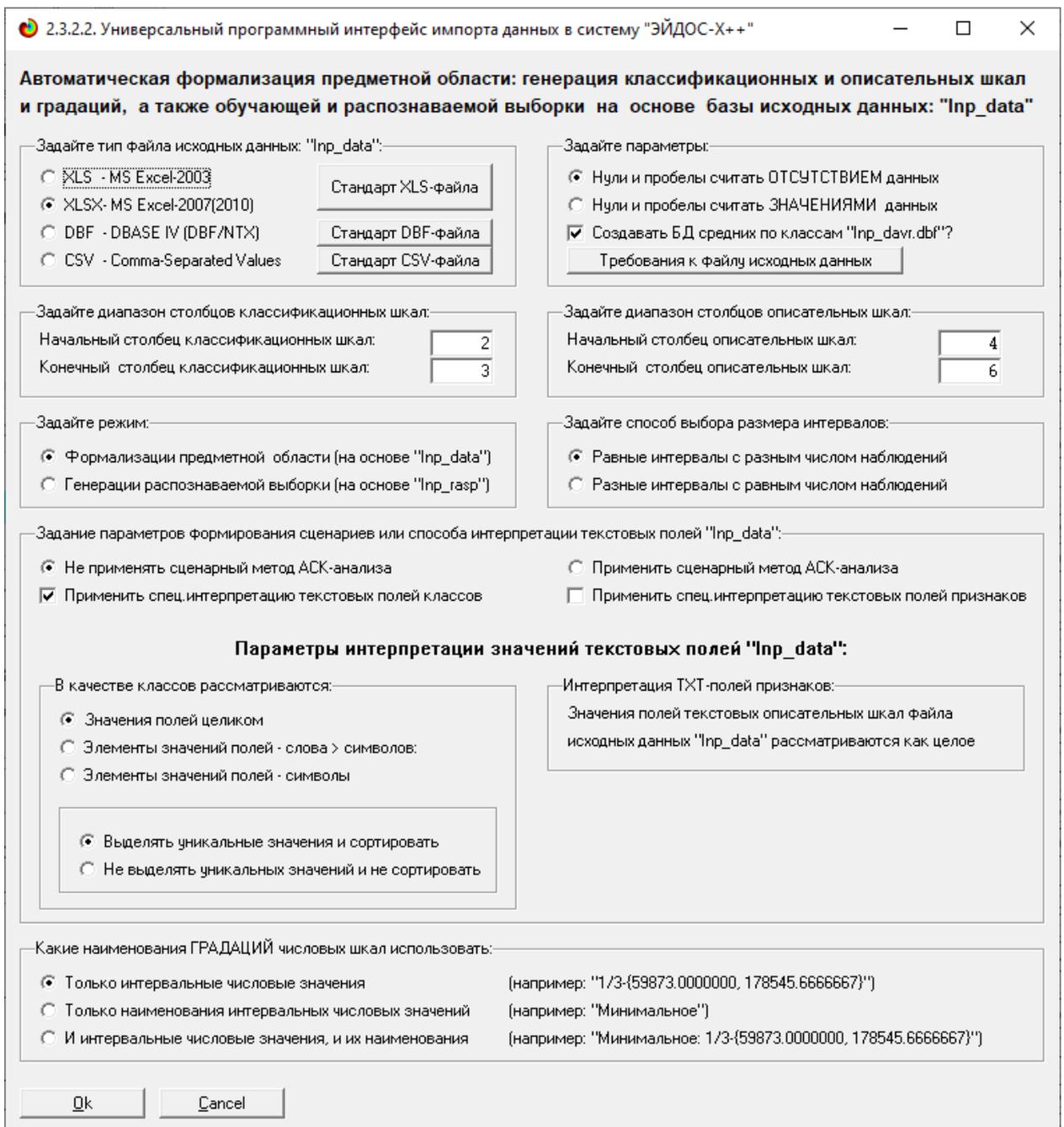


Рисунок 2 – Экранная форма универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2)

В экранной форме, приведенной на рисунке (Рисунок 3), необходимо задать следующие настройки:

– «Задайте тип файла исходных данных «Inp\_data»: «XLSX – MS Excel-2007(2010)»;

– «Задайте диапазон столбцов классификационных шкал»: «Начальный столбец классификационных шкал» – 2;

– «Конечный столбец классификационных шкал» – 3 (второй столбец в таблице);

– «Задайте диапазон столбцов описательных шкал»: «Начальный столбец описательных шкал» – 4, «Конечный столбец описательных шкал» – 6;

– «Задание параметров формирования сценариев или способа интерпретации текстовых полей»: «Не применять сценарный метод АСК-анализа», «Применить спец. интерпретацию текстовых полей классов».

После заполнения необходимых данных, нажать кнопку «Ок». Через некоторое время появляется окно внутреннего калькулятора (Рисунок 4). Экранная форма позволяет оценить, сколько текстовых и числовых классификационных и описательных шкал система обнаружила при заданных параметрах и сколько в них обнаружено градаций.

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс. шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	1	3	3,00	1	10	10,00
Текстовые	1	7	7,00	2	813	406,50
ВСЕГО:	2	10	5,00	3	823	274,33

Задание в диалоге размерности модели

Суммарное количество градаций классификационных и описательных шкал: [10 x 823]

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах:  В описательных шкалах:

Пересчитать шкалы и градации Выйти на создание модели

Рисунок 3 – Экранная форма внутреннего калькулятора режима 2.3.2.2

Обратим внимание на то, что заданы адаптивные интервалы, учитывающее неравномерность распределения данных по диапазону значений, что важно при относительно небольшом числе наблюдений. Если бы интервалы были заданы равными по величине, то в них бы учитывалось сильно отличающееся число наблюдений, а в некоторых интервалах их бы могло не оказаться вовсе.

Если обнаружены шкалы числового типа, то появляется возможность задать количество интервальных числовых значений в этих шкалах. Это делается отдельно для классификационных и описательных шкал, таким образом, число интервальных числовых значений в классификационных и описательных шкалах может отличаться. Если это число изменяется, то необходимо кликнуть полевой кнопке, а затем уже выходить на создание модели. В результате выполнения данного режима формируются классификационные, описательные шкалы, градации и обучающая выборка.

После нажатия на кнопку «Выйти на создание модели» запускается процесс импорта данных из внешней БД «Inp\_data.xlsx» в систему «Эйдос» (Рисунок 5), а также прогноз времени завершения этого процесса.

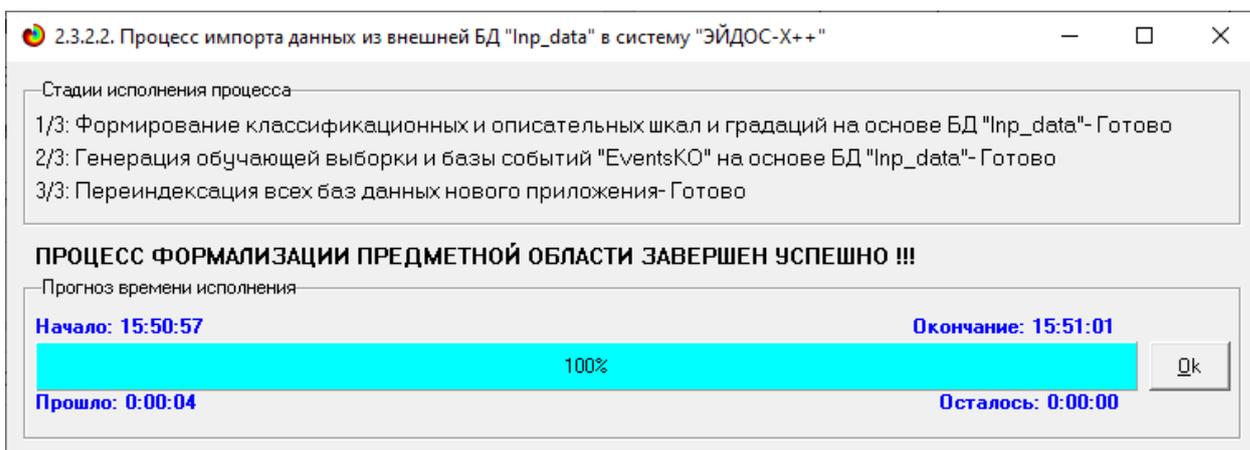


Рисунок 4 – Экранная форма импорта данных режима 2.3.2.2

В результате работы режима сформировано 1 классификационная шкала с суммарным количеством градаций (классов): 2 (Таблица 1) и 7 описательных шкал с суммарным числом градаций: 67 (Таблица 2).

С использованием классификационных и описательных шкал и градаций исходные данные (Рисунок 2) были закодированы и в результате получена обучающая выборка. Для просмотра классификационных шкал и градаций используется режим 2.1, для просмотра описательных шкал и градаций используется режим 2.2

Таблица 1 – Классификационные шкалы и градации

Код шкалы	Наименование классификационной	Код градации	Наименование градации
-----------	--------------------------------	--------------	-----------------------

	<b>шкалы</b>		
1	DAY_OF_WEEK	1	Friday
		2	Monday
		3	Saturday
		4	Sunday
		5	Thursday
		6	Tuesday
		7	Wednesday
2	HOUR	8	1/3-{1.0000000, 8.3333333}
		9	2/3-{8.3333333, 15.6666667}
		10	3/3-{15.6666667, 23.0000000}

Таблица 2 – Описательные шкалы и градации

<b>Код шкалы</b>	<b>Наименование классификационной шкалы</b>	<b>Код градации</b>	<b>Наименование градации</b>
1	MONTH	1	1/10-{5.0000000, 5.7000000}
		2	2/10-{5.7000000, 6.4000000}
		3	3/10-{6.4000000, 7.1000000}
		4	4/10-{7.1000000, 7.8000000}
		5	5/10-{7.8000000, 8.5000000}
		6	6/10-{8.5000000, 9.2000000}
		7	7/10-{9.2000000, 9.9000000}
2	OFFENSE_DESCRIPTION	8	8/10-{9.9000000, 10.6000000}
		9	9/10-{10.6000000, 11.3000000}
		10	10/10-{11.3000000, 12.0000000}
		11	4/10-{7.6000000,9.8000000}
		12	5/10-{9.8000000,12.0000000}
		13	6/10-{12.0000000,14.2000000}
		14	7/10-{14.2000000,16.4000000}
		15	8/10-{16.4000000,18.6000000}
		16	9/10-{18.6000000,20.8000000}
		17	10/10-{20.8000000,23.0000000}
3	STREET	18	1/10-{19.5000000, 19.9908333}
		19	2/10-{19.9908333, 20.4816667}
		20	3/10-{20.4816667, 20.9725000}
		21	4/10-{20.9725000, 21.4633333}
		22	5/10-{21.4633333, 21.9541667}
		23	6/10-{21.9541667, 22.4450000}
		24	7/10-{22.4450000, 22.9358333}
		25	8/10-{22.9358333, 23.4266666}
		26	9/10-{23.4266666, 23.9175000}

Код шкалы	Наименование классификационной шкалы	Код градации	Наименование градации
		27	10/10-{23.9175000, 24.4083333}

Обучающая выборка, по сути, представляет собой нормализованную с помощью классификационных и описательных шкал и градаций базу исходных данных. 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 (с учетом нелинейности системы [5]). Это делает исходные данные готовыми для обработки в программной системе и выполнения следующего этапа АСК-анализа: синтеза и верификации модели.

### **1.3. Синтез и верификация статистических и интеллектуальных моделей**

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) моделей осуществляется в режиме 3.5 системы «Эйдос» (Рисунок 6).

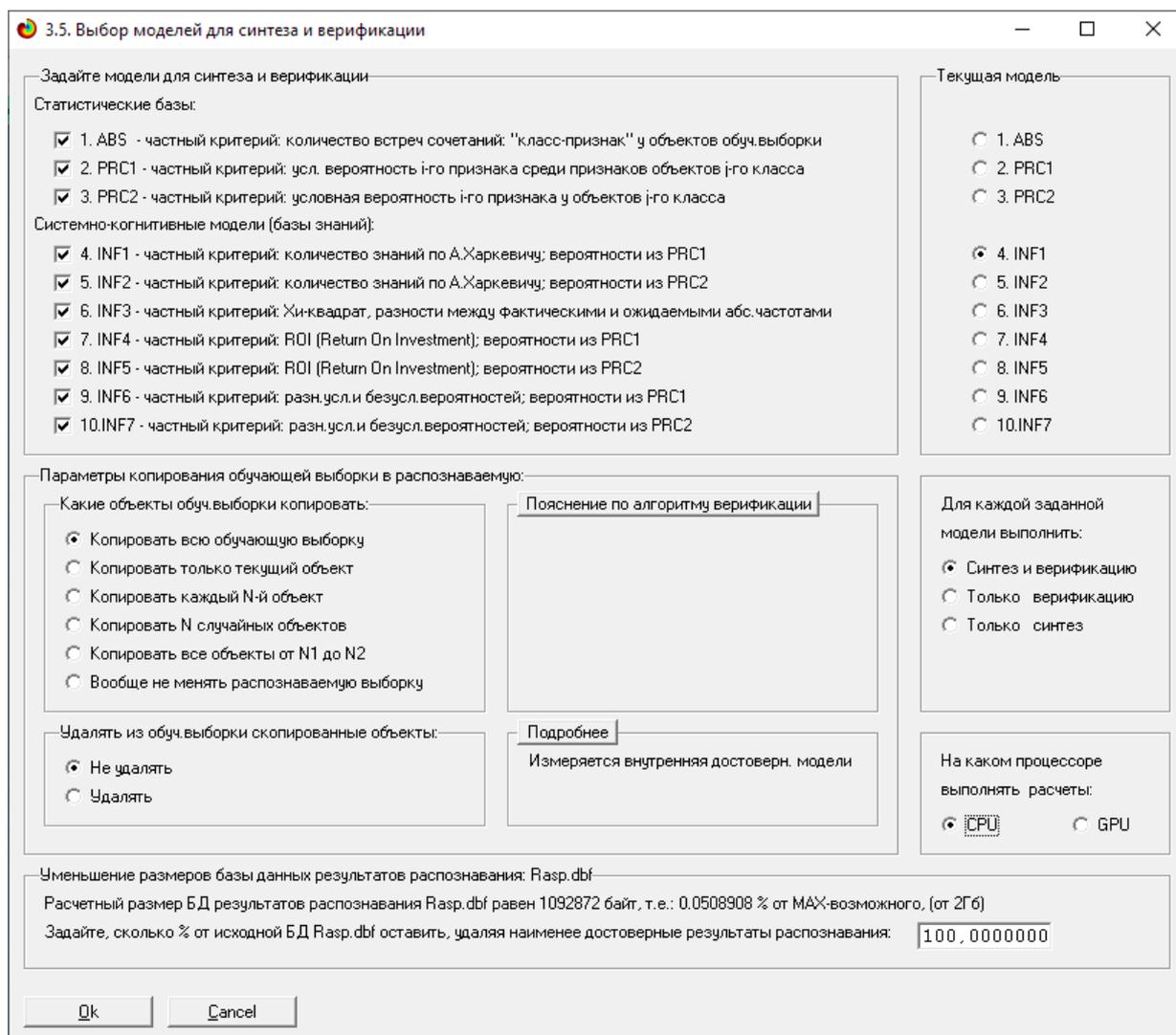


Рисунок 5 – Экранная форма выбора моделей для синтеза и верификации режима 3.5

В данном режиме имеется много различных методов верификации моделей, но в данной работе целесообразно использовать параметры по умолчанию. Необходимо обратить внимание на то, что для проведения вычислений можно использовать процессор GPU, если задать необходимое значение в опции: «Расчеты проводить на графическом процессор (GPU)».

Необходимо отметить, что синтез и верификация всех 10 моделей на ресурсах CPU заняли 20 минут 57 секунд (Рисунок 7), так как обрабатывается большой объем данных (более 1500 тысяч строк). При этом обработка той же выборки на ресурсах GPU заняла вдвое меньше времени – 10 минут 19 секунд. Таким образом, неграфические вычисления на графических процессорах видеокарты делает возможным обработку реальных текстов за разумное время.

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.

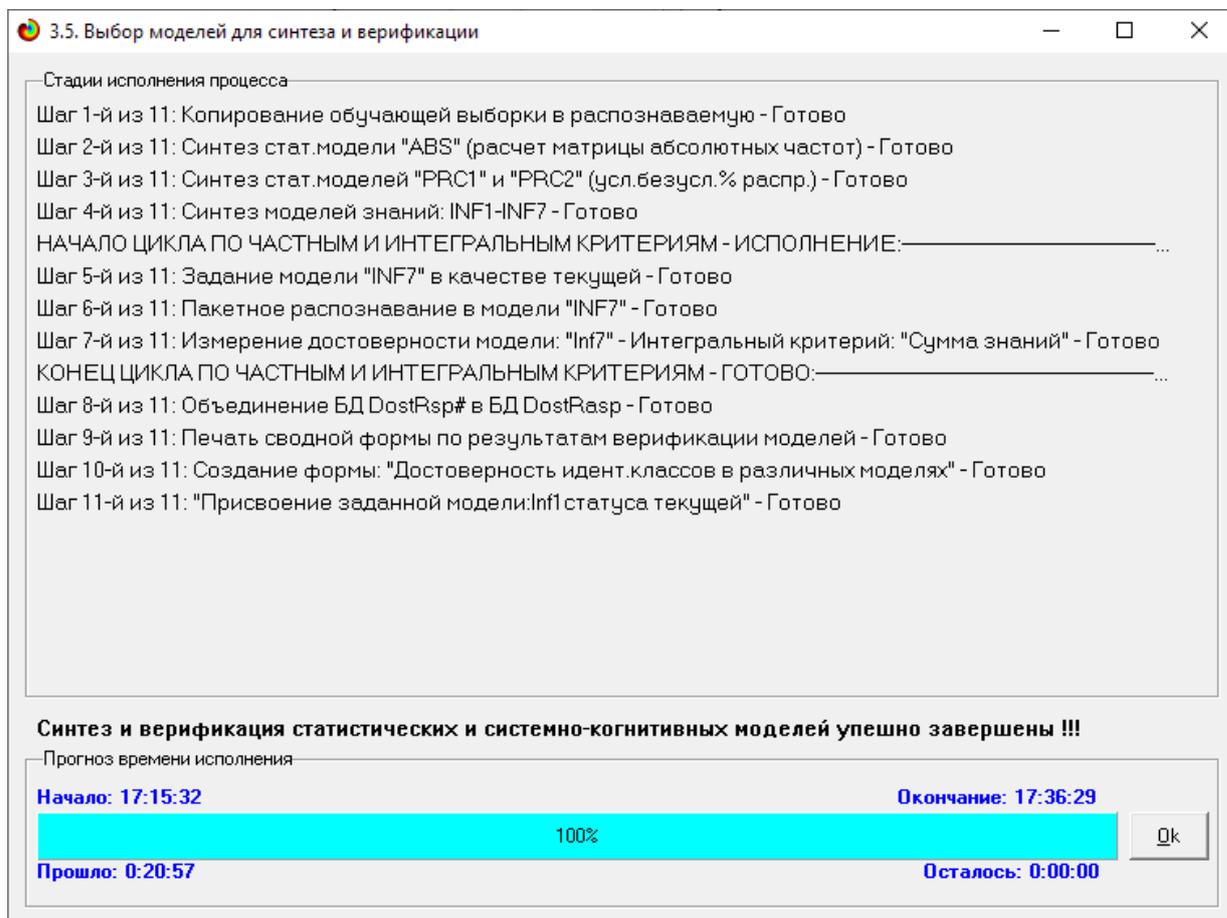


Рисунок 6 – Экранная форма с результатами работы режима 3.5

Фрагменты самих созданных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках (Рисунок 8, Рисунок 9, Рисунок 10).

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

Код признака	Наименование описательной шкалы и градации	1. DAY_OF_WEEK FRIDAY	2. DAY_OF_WEEK MONDAY	3. DAY_OF_WEEK SATURDAY	4. DAY_OF_WEEK SUNDAY	5. DAY_OF_WEEK THURSDAY	6. DAY_OF_WEEK TUESDAY	7. DAY_OF_WEEK WEDNESDAY	8. HOUR 1/3 (1.0, 8.3)	9. HOUR 2/3 (8.3, 15.7)	10. HOUR 3/3 (15.7, 23.0)	Сумма
1	MONTH-1/10-(5.000000, 5.700000)					1		2		1	2	6
2	MONTH-2/10-(5.700000, 6.400000)	1					1		1			4
3	MONTH-3/10-(6.400000, 7.100000)	1	1			1				4		8
4	MONTH-4/10-(7.100000, 7.800000)											
5	MONTH-5/10-(7.800000, 8.500000)	281	17	9	11	262	79	283	138	330	434	1844
6	MONTH-6/10-(8.500000, 9.200000)		176	264	227				148	224	263	1302
7	MONTH-7/10-(9.200000, 9.900000)											
8	MONTH-8/10-(9.900000, 10.600000)							1			1	2
9	MONTH-9/10-(10.600000, 11.300000)							1				1
10	MONTH-10/10-(11.300000, 12.000000)	1									1	2
11	OFFENSE_DESCRIPTION-AFFRAY				1					1		2
12	OFFENSE_DESCRIPTION-ANIMAL CONTROL - DOG...	2	1		1	2			2	3	1	12
13	OFFENSE_DESCRIPTION-ANIMAL INCIDENTS			1							1	2
14	OFFENSE_DESCRIPTION-ANNOYING AND ACCOSTI...	1									1	2
15	OFFENSE_DESCRIPTION-ARSON		1						1			2
16	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVATED ...	5	4	6	6	4	2	3	11	9	10	60
17	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVATED ...	4	3	1	3	1			2	4	4	22
18	OFFENSE_DESCRIPTION-ASSAULT - SIMPLE	1	1		3				3	1	2	12
19	OFFENSE_DESCRIPTION-ASSAULT SIMPLE - BATTE...	12	10	14	16	11	4	9	14	19	39	148
20	OFFENSE_DESCRIPTION-AUTO THEFT - MOTORCY...	2		1	3	2			1	2	2	16
21	OFFENSE_DESCRIPTION-AUTO THEFT	3	2	6	1	2		2	2	2	6	36
22	OFFENSE_DESCRIPTION-BALLISTICS EVIDENC/F...	1	1	1				1	1		4	10
23	OFFENSE_DESCRIPTION-BOMB THREAT				1		1				2	4

Рисунок 7 – Матрица абсолютных частот (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. DAY_OF_WEEK FRIDAY	2. DAY_OF_WEEK MONDAY	3. DAY_OF_WEEK SATURDAY	4. DAY_OF_WEEK SUNDAY	5. DAY_OF_WEEK THURSDAY	6. DAY_OF_WEEK TUESDAY	7. DAY_OF_WEEK WEDNESDAY	8. HOUR 1/3 (1.0, 8.3)	9. HOUR 2/3 (8.3, 15.7)	10. HOUR 3/3 (15.7, 23.0)	Сумма
1	MONTH-1/10-(5.000000, 5.700000)					0.246		0.470		-0.022	0.150	
2	MONTH-2/10-(5.700000, 6.400000)	0.371				0.393	0.829		0.368			
3	MONTH-3/10-(6.400000, 7.100000)	0.119	0.263			0.141		0.113		0.378		
4	MONTH-4/10-(7.100000, 7.800000)											
5	MONTH-5/10-(7.800000, 8.500000)	0.191	-0.684	-1.039	-0.916	0.188	0.188	0.188	-0.070	0.004	0.023	
6	MONTH-6/10-(8.500000, 9.200000)		0.291	0.315	0.310				0.082	-0.010	-0.032	
7	MONTH-7/10-(9.200000, 9.900000)											
8	MONTH-8/10-(9.900000, 10.600000)							0.617			0.297	
9	MONTH-9/10-(10.600000, 11.300000)	0.623						0.869				
10	MONTH-10/10-(11.300000, 12.000000)					0.693					0.378	
11	OFFENSE_DESCRIPTION-AFFRAY					0.042	0.246		0.221	0.126	-0.354	
12	OFFENSE_DESCRIPTION-ANIMAL CONTROL - D...	0.224	0.116								0.297	
13	OFFENSE_DESCRIPTION-ANIMAL INCIDENTS			0.642							0.297	
14	OFFENSE_DESCRIPTION-ANNOYING AND ACC...	0.623									0.297	
15	OFFENSE_DESCRIPTION-ARSON		0.766					0.620				
16	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVA...	-0.028	0.035	0.058	0.108	-0.087	0.097	-0.219	0.256	-0.060	-0.102	
17	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVA...	0.256	0.294	-0.229	0.221	-0.226			0.001	0.010	-0.070	
18	OFFENSE_DESCRIPTION-ASSAULT - SIMPLE	-0.028	0.116		0.441			-0.034	0.368	-0.273	-0.102	
19	OFFENSE_DESCRIPTION-ASSAULT SIMPLE - B...	-0.038	0.039	0.038	0.137	-0.048	0.021	-0.148	0.015	-0.116	0.065	
20	OFFENSE_DESCRIPTION-AUTO THEFT - MOTO...	0.119		-0.113	0.337	0.141		-0.138	0.116	-0.126	-0.059	
21	OFFENSE_DESCRIPTION-AUTO THEFT	-0.028	-0.032	0.243	-0.357	-0.153		0.283	-0.181	-0.178	-0.022	0.084
22	OFFENSE_DESCRIPTION-BALLISTICS EVIDENC...	0.038	0.182	0.058				0.496	0.032	0.035	0.216	
23	OFFENSE_DESCRIPTION-BOMB THREAT				0.441		0.829				0.297	

Рисунок 8 – Матрица информативностей INF1 (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. DAY_OF_WEEK FRIDAY	2. DAY_OF_WEEK MONDAY	3. DAY_OF_WEEK SATURDAY	4. DAY_OF_WEEK SUNDAY	5. DAY_OF_WEEK THURSDAY	6. DAY_OF_WEEK TUESDAY	7. DAY_OF_WEEK WEDNESDAY	8. HOUR 1/3 (1.0, 8.3)	9. HOUR 2/3 (8.3, 15.7)	10. HOUR 3/3 (15.7, 23.0)	Сумма
1	MONTH-1/10-(5.000000, 5.700000)	-0.540	-0.364	-0.512	-0.445	0.492	-0.153	1.451	-0.544	-0.061	0.676	
2	MONTH-2/10-(5.700000, 6.400000)	0.640	-0.243	-0.341	-0.297	0.661	0.898	-0.366	0.637	-0.707	-0.882	
3	MONTH-3/10-(6.400000, 7.100000)	0.280	0.515	-0.682	-0.593	0.322	-0.204	0.268	-0.726	2.595	-1.765	
4	MONTH-4/10-(7.100000, 7.800000)											
5	MONTH-5/10-(7.800000, 8.500000)	115.081	-94.793	-148.259	-125.789	105.725	31.960	114.326	-29.296	3.870	27.175	
6	MONTH-6/10-(8.500000, 9.200000)	-117.151	97.066	152.964	130.417	-110.341	-33.214	-119.096	29.876	-6.272	-24.249	
7	MONTH-7/10-(9.200000, 9.900000)											
8	MONTH-8/10-(9.900000, 10.600000)	-0.180	-0.121	-0.171	-0.148	-0.169	-0.051	0.817	-0.181	-0.354	0.559	
9	MONTH-9/10-(10.600000, 11.300000)	-0.090	-0.061	-0.085	-0.074	-0.085	-0.026	0.909	-0.091	-0.177	-0.221	
10	MONTH-10/10-(11.300000, 12.000000)	0.820	-0.121	-0.171	-0.148	-0.169	-0.051	-0.183	-0.181	-0.354	0.559	
11	OFFENSE_DESCRIPTION-AFFRAY	-0.180	-0.121	-0.171	0.852	-0.169	-0.051	-0.183	-0.181	0.646	-0.441	
12	OFFENSE_DESCRIPTION-ANIMAL CONTROL - D...	0.920	0.272	-1.023	0.110	0.983	-0.306	-1.098	0.911	0.878	-1.647	
13	OFFENSE_DESCRIPTION-ANIMAL INCIDENTS	-0.180	-0.121	0.829	-0.148	-0.169	-0.051	-0.183	-0.181	-0.354	0.559	
14	OFFENSE_DESCRIPTION-ANNOYING AND ACC...	0.820	-0.121	-0.171	-0.148	-0.169	-0.051	-0.183	-0.181	-0.354	0.559	
15	OFFENSE_DESCRIPTION-ARSON	-0.180	0.879	-0.171	-0.148	-0.169	-0.051	-0.183	0.819	-0.354	-0.441	
16	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVA...	-0.399	0.362	0.883	1.549	-1.085	0.469	-2.488	5.557	-1.612	-3.237	
17	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVA...	2.020	1.666	-0.876	1.368	-0.864	-0.561	-2.012	0.004	0.109	-0.854	
18	OFFENSE_DESCRIPTION-ASSAULT - SIMPLE	-0.080	0.272	-1.023	2.110	-1.017	-0.306	-0.098	1.911	-1.122	-0.647	
19	OFFENSE_DESCRIPTION-ASSAULT SIMPLE - B...	-1.317	1.027	1.378	5.021	-1.543	0.225	-4.538	0.573	-7.175	6.348	
20	OFFENSE_DESCRIPTION-AUTO THEFT - MOTO...	0.560	-0.970	-0.364	1.813	0.644	-0.408	-0.464	0.548	-0.830	-0.530	
21	OFFENSE_DESCRIPTION-AUTO THEFT	-0.239	-0.183	2.930	-1.671	-1.051	1.082	-1.293	-1.266	-0.367	2.058	
22	OFFENSE_DESCRIPTION-BALLISTICS EVIDENC...	0.100	0.394	0.147	-0.742	-0.847	0.745	0.085	0.093	-1.769	1.794	
23	OFFENSE_DESCRIPTION-BOMB THREAT	-0.360	-0.243	-0.341	0.703	-0.339	0.898	-0.366	-0.363	-0.707	1.118	

Рисунок 9 – Модель INF3 (фрагмент)

Отметим, что в АСК-анализе и СК-моделях степень выраженности различных свойств объектов наблюдения рассматривается с единственной точки зрения: с точки зрения того, какое количество информации содержится в них о том, к каким обобщающим категориям (классам) будут принадлежать или не принадлежать эти объекты. Поэтому не играет никакой роли в каких единицах измерения измеряются те или иные свойства объектов наблюдения, а также в каких единицах измеряются результаты влияния этих свойств, натуральных, в процентах или стоимостных [6]. Это и есть решение проблемы сопоставимости в АСК-анализе и системе «Эйдос», отличающее их от других интеллектуальных технологий.

#### **1.4. Верификация статистических и системно-когнитивных моделей**

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по обобщенным образам классов и подсчета количества истинных положительных и отрицательных, а также ложных положительных и отрицательных решений по F-мере Ван Ризбергена, а также по критериям L1-L2-мерам проф. Е.В. Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры [7]. В режиме 3.4 системы «Эйдос» изучается достоверность каждой частной модели в соответствии с этими мерами достоверности.

Из рисунка (Рисунок 11) мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной является СК-модель INF1 с интегральным критерием «Семантический резонанс знаний» ( $F=0,959$  при максимуме 1), что является довольно хорошим результатом, по критерию L1 проф. Е.В. Луценко наиболее достоверной также является СК-модель INF3 с интегральным критерием «Сумма знаний» ( $L1=0,975$  при максимуме 1), что также является хорошим результатом.

3.4. Обобщ.форма по достов.моделей при разн.инт.крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F	V	P
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс частот с обр...	3169	3168	537	12514	1	0.202	1.000			
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс частот по признак...	3169	3169	245	12806		0.198	1.000			
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	3169	3168	537	12514	1	0.202	1.000			
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по приз...	3169	3169	245	12806		0.198	1.000			
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	3169	3168	537	12514	1	0.202	1.000			
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по приз...	3169	3169	245	12806		0.198	1.000			
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	3169	2062	11175	1876	1107	0.524	0.651			
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	3169	2865	6662	6389	304	0.310	0.904			
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	3169	2071	11014	2037	1098	0.504	0.654			
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	3169	2866	6635	6416	303	0.309	0.904			
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	3169	2514	7311	5740	655	0.305	0.793			
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	3169	2514	7311	5740	655	0.305	0.793			
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	3169	1772	12546	505	1397	0.778	0.559			
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	3169	2943	5920	7131	226	0.292	0.929			
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс зна...	3169	1771	12546	505	1398	0.778	0.559			
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	3169	2947	5924	7127	222	0.293	0.930			
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Семантический резонанс зна...	3169	2502	6677	6374	667	0.282	0.790			
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Сумма знаний	3169	2514	6304	6747	655	0.271	0.793			
10. INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Семантический резонанс зна...	3169	2517	6647	6404	652	0.282	0.794			
10. INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Сумма знаний	3169	2530	6269	6782	639	0.272	0.798			

Помощь по мерам достоверности | Помощь по частотным распределениям | TP,TN,FP,FN | (TP-FP), (TN-FN) | (T-F)/(T+F)\*100 | Задать интервал сглаживания

Рисунок 10 – Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1- и L2-критериям проф. Е.В. Луценко (фрагмент)

Хотелось бы обратить внимание, что статистические модели, как правило, дают более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую. Этим и оправдано применение моделей знаний и интеллектуальных технологий.

Это позволяет сделать обоснованные выводы о том, что:

- существует достаточно сильная причинно-следственная зависимость между данными датчиков и присутствием человека в помещении;
- эта причинно-следственная зависимость адекватно отражена в системно-когнитивных моделях INF1 и INF3;
- так как L2-мера проф. Е.В. Луценко для модели INF1 составляет 0,620, а для INF3 – 0,787, то предпочтение отдаем модели INF3.

На рисунке (Рисунок 12) приведено частотное распределения числа истинных и ложных положительных и отрицательных решений по результатам оценки данных датчиков в СК-модели INF3 по данным обучающей выборки.

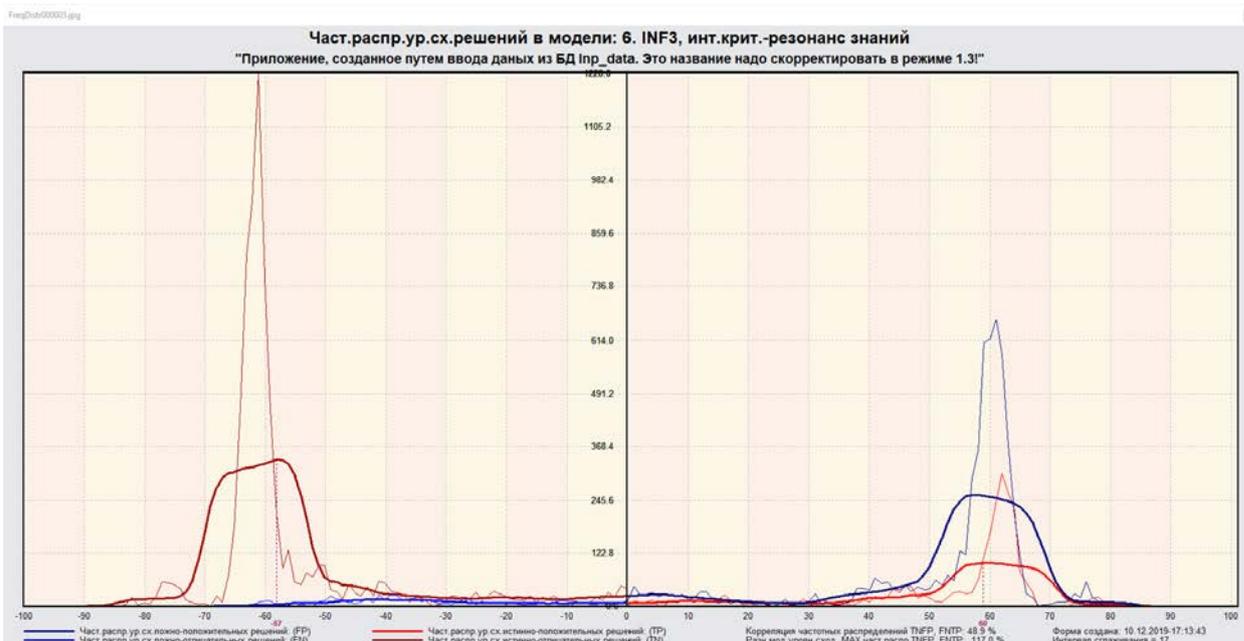


Рисунок 11 – Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели INF3

На рисунке (Рисунок 13) изображены частотные распределения разностей долей разностей уровней сходимости решений в СК-модели INF3.

Исходя из данных на рисунках (Рисунок 12, Рисунок 13) можно сделать следующие выводы:

- левое и правое распределения практически равны по амплитуде. Левое распределение включает в себя как истинно-отрицательные, так и ложно-положительные решения, а правое, включает ложные отрицательные и истинно-положительные решения;

- сдвиг этих распределений относительно друг друга и другие различия между ними позволяют решать задачу присутствия человека в помещении на основании данных датчиков, а также другие задачи;

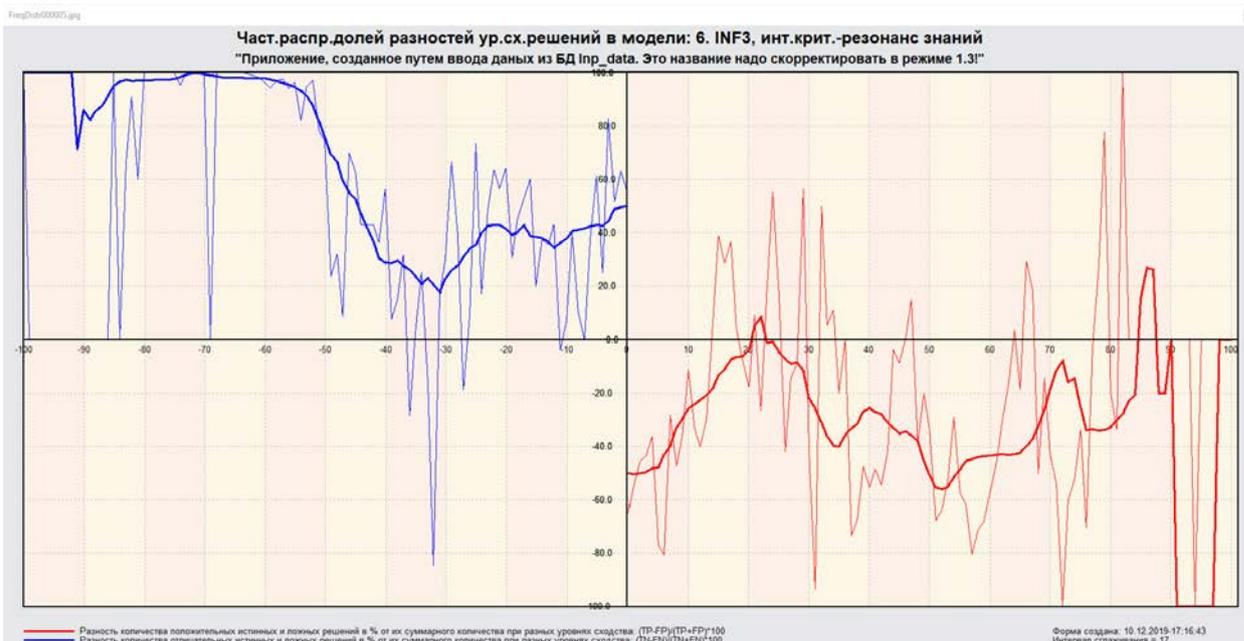


Рисунок 12 – Частотные распределения разностей долей разностей уровней сходимости решений в СК-модели INF3

– видно, что для отрицательных решений количество истинных решений всегда значительно превосходит количество ложных решений, причем при уровнях различия больше примерно 17% ложные отрицательные решения вообще практически отсутствуют;

– также видно, что для положительных решений при уровнях сходства от 0% до примерно 62% есть и истинные и ложные положительные решения, но число истинных решений больше числа ложных и доля истинных решений возрастает при увеличении уровня сходства. При уровнях сходства выше 62% ложно-положительных решений практически не встречается;

– модель INF3 немного лучше определяет непринадлежность объекта к классу, чем принадлежность (процент правильной не идентификации выше).

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями, совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются

вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически (Рисунок 14).

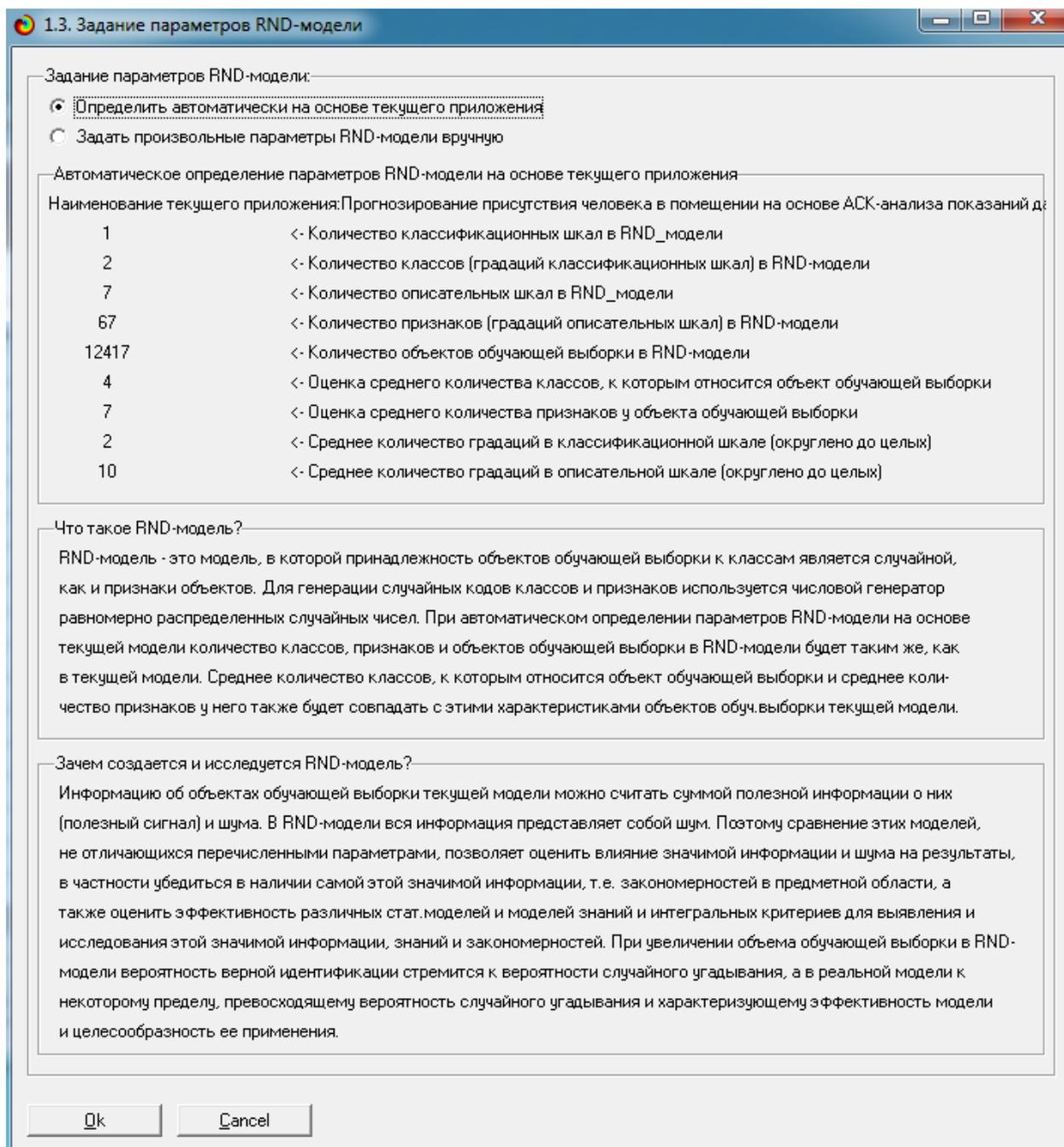


Рисунок 13 – Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке (Рисунок 16) показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3. Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта

моделирования и случайной модели (Рисунок 15) объясняется тем, что в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны. Если же такой информации в модели нет, то и распределение получается типа, приведенного на рисунке (Рисунок 16).

4.1.3.6. Обзор форма по достов. моделей при разн. крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полезна модель	Ф-мера Ван Ризбергера	Сумма модул. уровней сход. истинно-полож. решений (STR)	Сумма модул. уровней сход. истинно-отриц. решений (STN)	Сумма уровней ложно-полож. решений	Сумма уровней ложно-отриц. решений
1. ABS - частный критерий: количество встреч совпадений: "Улос...	Корреляция абс. частот с обр...	3169	3168	537	12514	1	0.202	1.000	0.336	2128.350	2.924	5351	
1. ABS - частный критерий: количество встреч совпадений: "Улос...	Сумма абс. частот по призна...	3169	3169	245	12806		0.198	1.000	0.331	1734.590		3668	
2. PRC1 - частный критерий: усл. вероятность н-го признака сред...	Корреляция усл. частот с о...	3169	3168	537	12514	1	0.202	1.000	0.336	2128.350	2.924	5351	
2. PRC1 - частный критерий: усл. вероятность н-го признака сред...	Сумма усл. частот по призна...	3169	3169	245	12806		0.198	1.000	0.331	2150.463		5580	
3. PRC2 - частный критерий: условная вероятность н-го признака...	Корреляция усл. частот с о...	3169	3168	537	12514	1	0.202	1.000	0.336	2128.335	2.924	5351	
3. PRC2 - частный критерий: условная вероятность н-го признака...	Сумма усл. частот по призна...	3169	3169	245	12806		0.198	1.000	0.331	2180.620		5647	
4. INF1 - частный критерий: количество знаний по А\Харкивнен; в...	Семантический резонанс: зна...	3169	2062	11175	1076	1107	0.524	0.651	0.580	378.203	2699.923	140	
4. INF1 - частный критерий: количество знаний по А\Харкивнен; в...	Сумма знаний	3169	2865	6662	6989	304	0.310	0.904	0.461	650.395	1286.101	602	
5. INF2 - частный критерий: количество знаний по А\Харкивнен; в...	Семантический резонанс: зна...	3169	2071	11014	2037	1098	0.504	0.654	0.569	375.223	2696.509	147	
5. INF2 - частный критерий: количество знаний по А\Харкивнен; в...	Сумма знаний	3169	2866	6635	6416	303	0.309	0.904	0.460	647.378	1286.757	604	
6. INF3 - частный критерий: Уинкварт, разности между фактич...	Семантический резонанс: зна...	3169	2514	7311	5740	655	0.305	0.793	0.440	1336.959	4078.123	3092	
6. INF3 - частный критерий: Уинкварт, разности между фактич...	Сумма знаний	3169	2514	7311	5740	655	0.305	0.793	0.440	1194.006	3712.224	2605	
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс: зна...	3169	1772	12546	505	1397	0.778	0.559	0.651	205.557	1242.842	17	
7. INF4 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	3169	2943	5920	7131	226	0.292	0.929	0.444	197.004	71.720	144	
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Семантический резонанс: зна...	3169	1771	12546	505	1398	0.778	0.559	0.651	205.570	1244.451	17	
8. INF5 - частный критерий: ROI (Return On Investment); вероятн...	Сумма знаний	3169	2947	5924	7127	222	0.293	0.930	0.444	194.856	71.285	141	
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс: зна...	3169	2502	6677	6374	667	0.282	0.790	0.415	1446.545	2278.240	3530	
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	3169	2514	6304	6747	655	0.271	0.793	0.405	1222.979	2489.326	3062	
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс: зна...	3169	2517	6647	6404	652	0.282	0.794	0.416	1442.513	2284.646	3529	
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	3169	2530	6269	6782	639	0.272	0.790	0.405	1245.903	2537.399	3138	

Рисунок 14 – Экранная форма с информацией о достоверности RND-моделей по F-критерию Ван Ризбергера и L1- и L2-критериям проф. Е.В. Луценко (фрагмент)

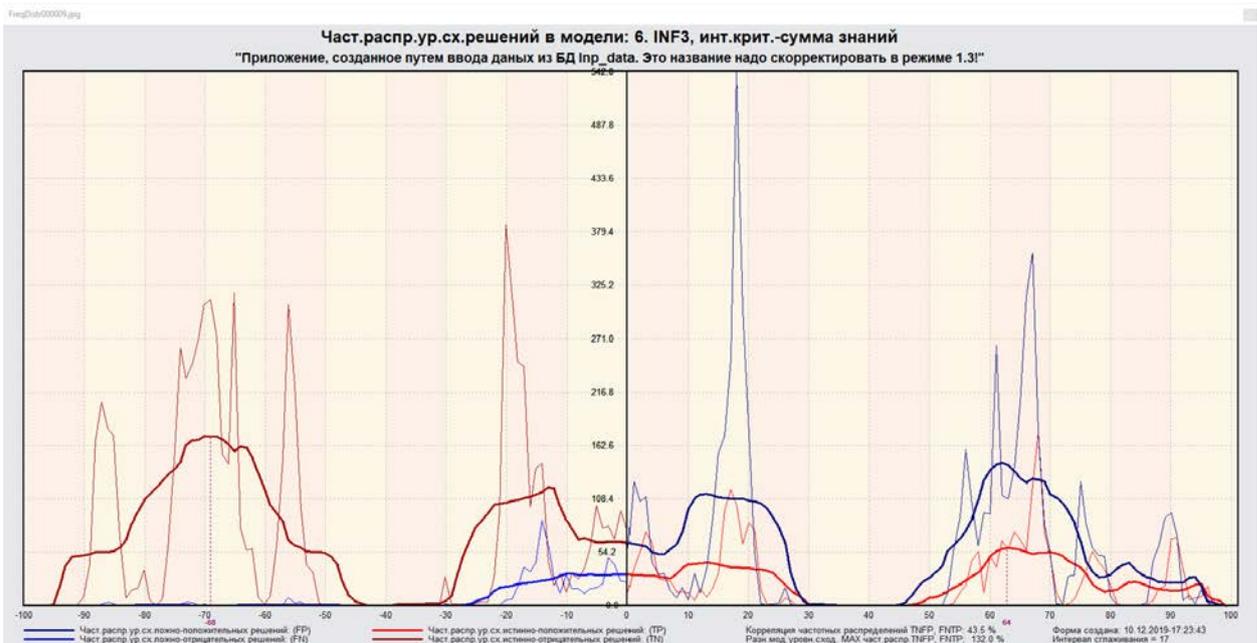


Рисунок 15 – Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в RND-модели INF3

На основе его сравнения с рисунком (Рисунок 12) можно сделать следующие выводы:

– RND-модель дает ложно-положительные и ложно-отрицательные прогнозы практически на всем диапазоне уровней сходства-различия;

– достоверность лучшей модели INF3, отражающей реальный объект моделирования, примерно на 50% выше, чем аналогичной случайной модели;

– различие между достоверностью статистических моделей и моделей знаний, созданных на основе случайной выборки, значительно меньше, чем у моделей, отражающих реальный объект моделирования;

– в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны, причем примерно 1/3 достоверности обусловлена отражением в реальных моделях закономерностей предметной области, а 2/3 достоверности обусловлено наличием шума в исходных данных.

На основании этого можно предположить, что в исходных данных уровень сигнала о реальных причинно-следственных связях в моделируемой предметной области примерно в два раза ниже уровня шума [8].

## **2. Решение задач идентификации в наиболее достоверной модели**

### **2.1. Выбор наиболее достоверной модели и присвоение ей статуса текущей**

В соответствии со схемой обработки данных, информации и знаний в системе «Эйдос», присвоим СК-модели INF3 статус текущей модели. Для этого запустим режим 5.6 с параметрами, приведенными на экранной форме (Рисунок 17), результаты операции изображены на рисунке (Рисунок 18).

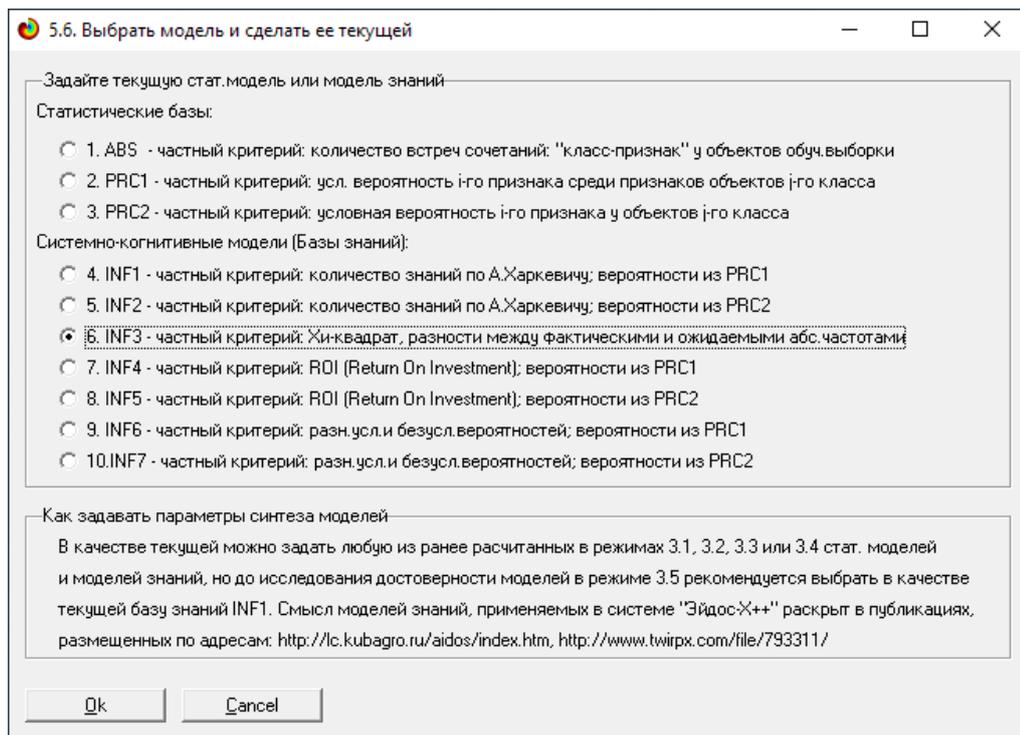


Рисунок 16 – Экранная форма придания наиболее достоверной по F-критерию СК-модели INF3 статуса текущей модели

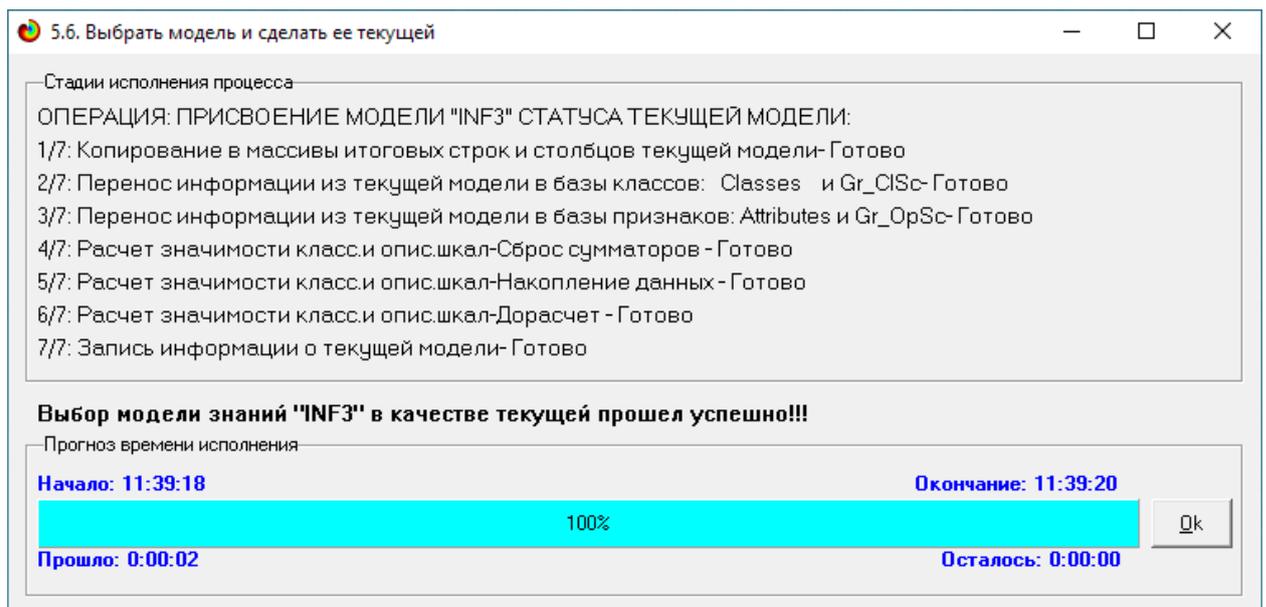


Рисунок 17 – Экранная формы результата присвоения статуса текущей модели

## 2.2. Прогнозирование (диагностика, классификация, распознавание, идентификация)

Решим задачу системной идентификации, т.е. определение присутствия человека на основе обучающей выборки в наиболее достоверной СК-модели INF3 (Рисунок 7).



На рисунках (Рисунок 20, Рисунок 21) приведены примеры прогнозов высокой и низкой достоверности классов, а также итоги распознавания классов объектов в модели INF3.

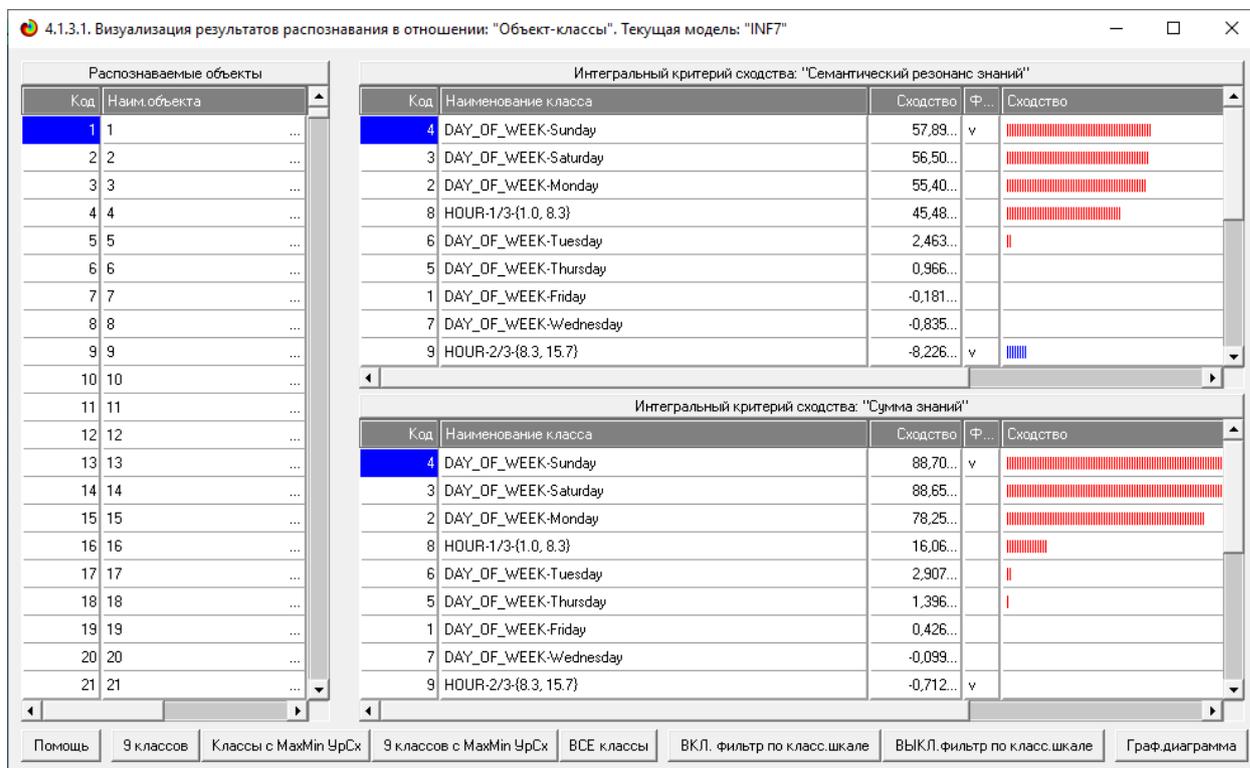


Рисунок 19 – Экранная форма визуализации результатов распознавания в отношении «Объект-классы» модели INF3 (режим 4.1.3.1)

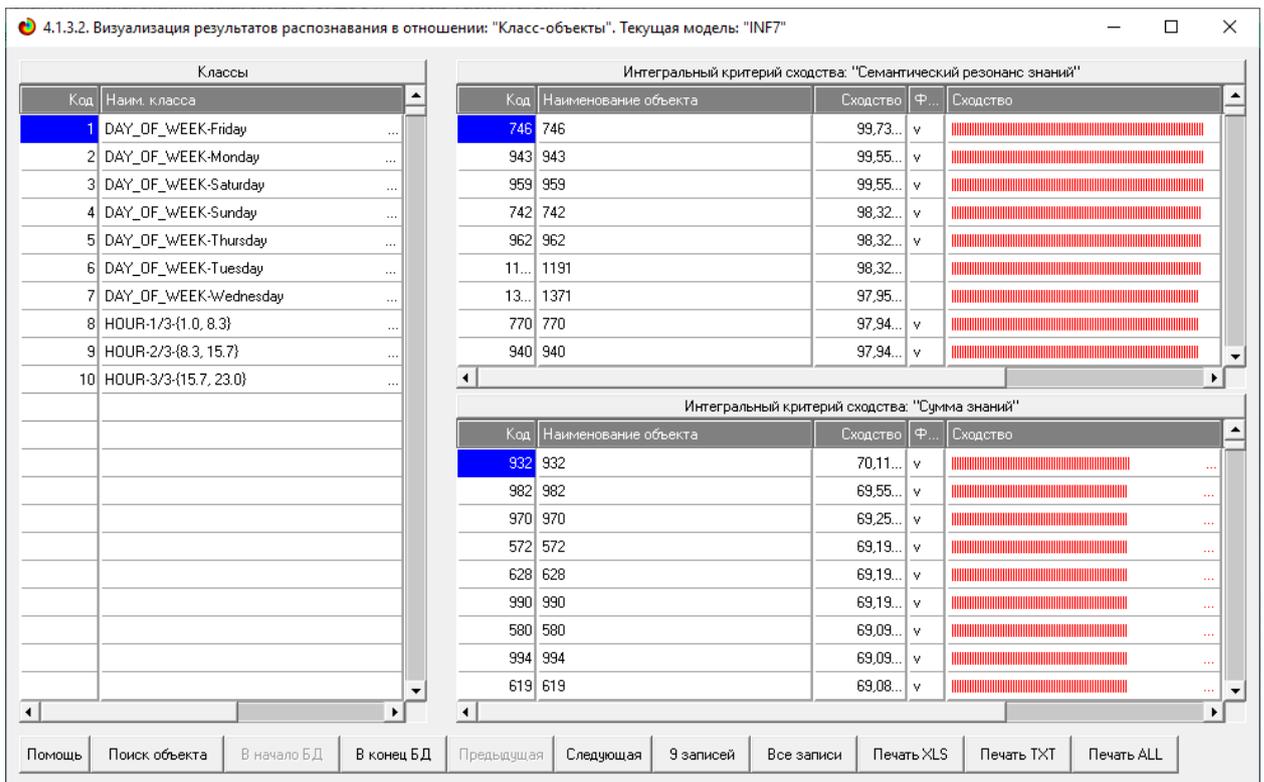


Рисунок 20 – Экранная форма визуализации результатов распознавания в отношении «Классы-объекты» модели INF3 (режим 4.1.3.2)

Символ «v» стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка (Рисунок 20) видно, что результаты идентификации являются хорошими, естественно при учете информации из рисунка (Рисунок 12) о том, что достоверные прогнозы в данной модели имеют уровень сходства выше 62%, т.е. по сути результаты с более низки уровнем сходства надо просто игнорировать.

### 2.3. Визуализация когнитивных функций

Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора на переходы объекта управления в будущие состояния, соответствующие классам.

Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации закономерностей и эмпирических законов. Разработка содержательной научной интерпретации

когнитивных функций представляет собой способ познания природы, общества и человека.

Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации; негативные, отражающие чему препятствуют система детерминации; средневзвешенные, отражающие совокупное влияние всех значений факторов на поведение объекта (причем в качестве весов наблюдений используется количество информации в значении аргумента о значениях функции) различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала.

Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию.

Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой «Эйдос». Необходимо отметить, что на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений, в частности, они могут быть и не дифференцируемые.

Применительно к задачам, рассматриваемым 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. В системе «Эйдос» за формирование когнитивных функций отвечает режим 4.5.

На рисунках (Рисунок 22, Рисунок 23) приведены примеры некоторых когнитивных функций, наглядно отражающих силу и направление влияния значений (т.е. степени выраженности) места преступления.

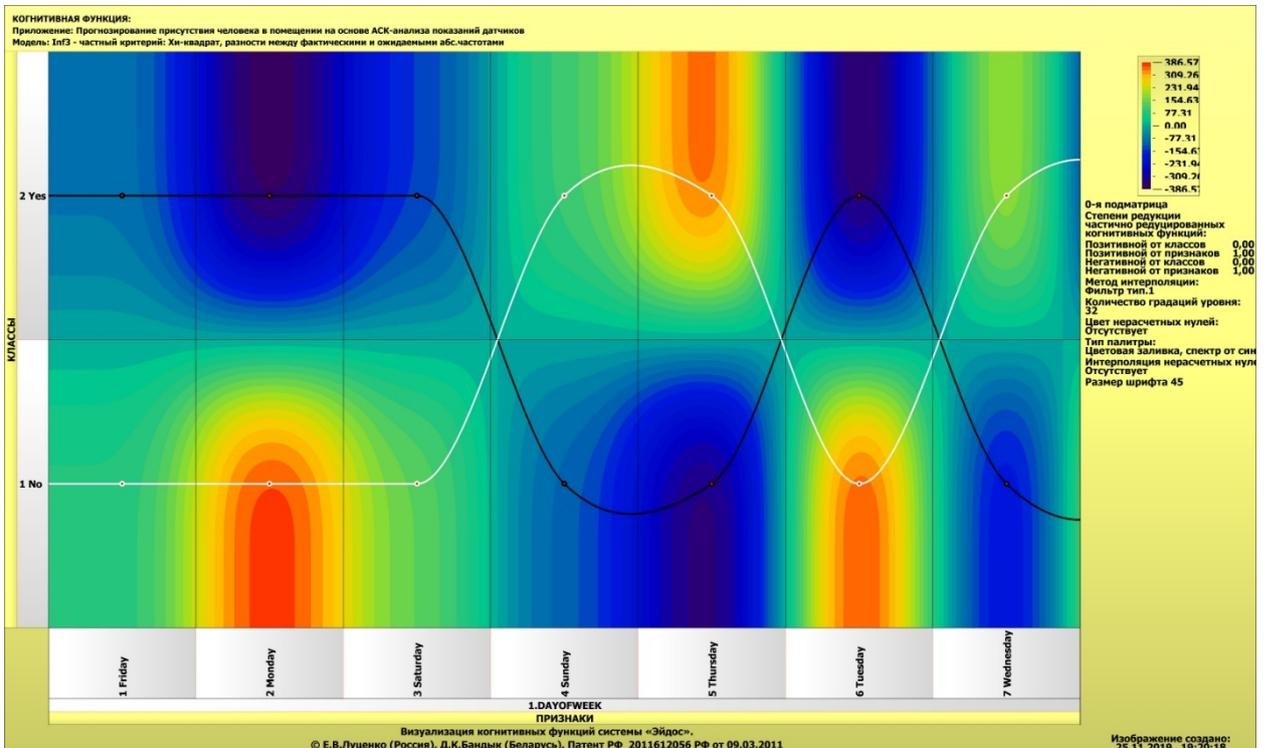


Рисунок 21 – Когнитивная функция, отражающая взаимосвязь места преступления и дня недели

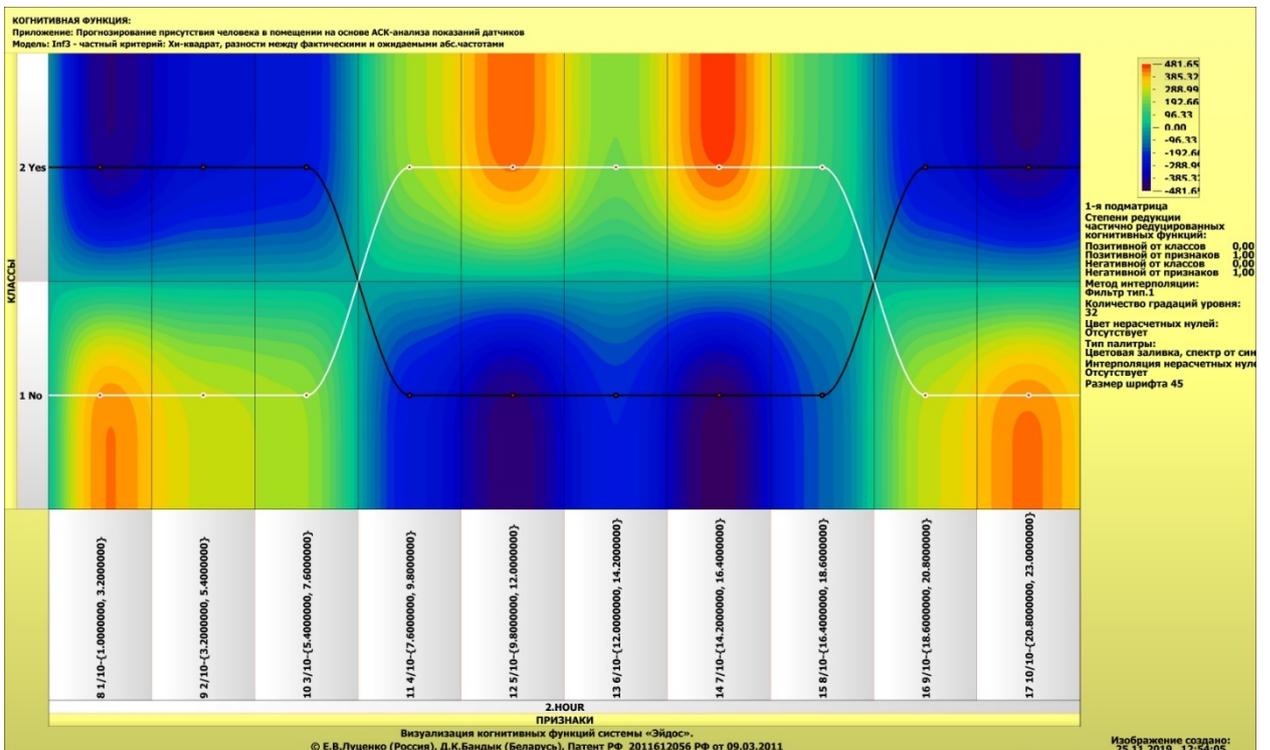


Рисунок 22 – Когнитивная функция, отражающая взаимосвязь места преступления и часа

В результате анализа полученных данных можно сделать следующие **ВЫВОДЫ:**

4.2.1. Информационные портреты классов

Инф.портрет класса:1 "DAY\_OF\_WEEK-Friday" в модели:6 "INF3"

Код	Наименование класса	Код	Наименование признака	Значимость
1	DAY_OF_WEEK-Friday	5	MONTH-5/10-(7.8000000, 8.5000000)	115.081
2	DAY_OF_WEEK-Monday	130	OFFENSE_DESCRIPTION-VANDALISM	5.074
3	DAY_OF_WEEK-Saturday	214	STREET-BOYLSTON ST	3.612
4	DAY_OF_WEEK-Sunday	449	STREET-HEATH ST	3.280
5	DAY_OF_WEEK-Thursday	104	OFFENSE_DESCRIPTION-PROPERTY - LOST	2.912
6	DAY_OF_WEEK-Tuesday	81	OFFENSE_DESCRIPTION-LIQUOR - DRINKING IN PUBLIC	2.740
7	DAY_OF_WEEK-Wednesday	190	STREET-BENNINGTON ST	2.740
8	HOUR-1/3-(1.0, 8.3)	164	STREET-ARCADIA ST	2.460
9	HOUR-2/3-(8.3, 15.7)	776	STREET-WASHINGTON ST	2.413
10	HOUR-3/3-(15.7, 23.0)	385	STREET-FIDELIS WAY	2.370
		83	OFFENSE_DESCRIPTION-M/V - LEAVING SCENE - PROPERTY DAMAGE	2.334
		209	STREET-BOSTON ST	2.280
		374	STREET-FANEUIL HALL MARKETPLACE	2.280
		80	OFFENSE_DESCRIPTION-LICENSE PREMISE VIOLATION	2.100
		17	OFFENSE_DESCRIPTION-ASSAULT - AGGRAVATED	2.020
		466	STREET-HUMBOLDT AVE	1.920
		679	STREET-SEAVER ST	1.740
		40	OFFENSE_DESCRIPTION-DRUGS - POSS CLASS B - COCAINE, ETC.	1.671
		105	OFFENSE_DESCRIPTION-PROPERTY - LOST THEN LOCATED	1.640
		172	STREET-ASHMONT ST	1.640
		277	STREET-CLAYTON ST	1.640
		313	STREET-DALE ST	1.640
		342	STREET-E FIRST ST	1.640

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 MS Excel Вкл.фильтр по фактору Выкл.фильтр по фактору Вписать в окно Показать ВСЕ

Рисунок 24 – Информационный портрет преступлений по дням недели

Согласно информационным портретам класса DAY\_OF\_WEEK (Рисунок 24) на определение наиболее опасного места в определенный день недели, мы видим, что признак под наименованием MONTH занимает наиболее опасную, лидирующую позицию .

## 2.4. Нелокальные нейроны и нейронная сеть

Для каждого технологического фактора в соответствии с предложенной моделью определяется величина и направление его влияния на осуществление всех желаемых и не желаемых хозяйственных ситуаций. Для каждой ситуации эта информация отображается в различных текстовых и графических формах, в частности, в форме нелокального нейрона (Рисунок

25, Рисунок 26). На данной диаграмме цвет линии означает знак связи (красный – положительная, синий – отрицательная), а толщина – ее модуль.

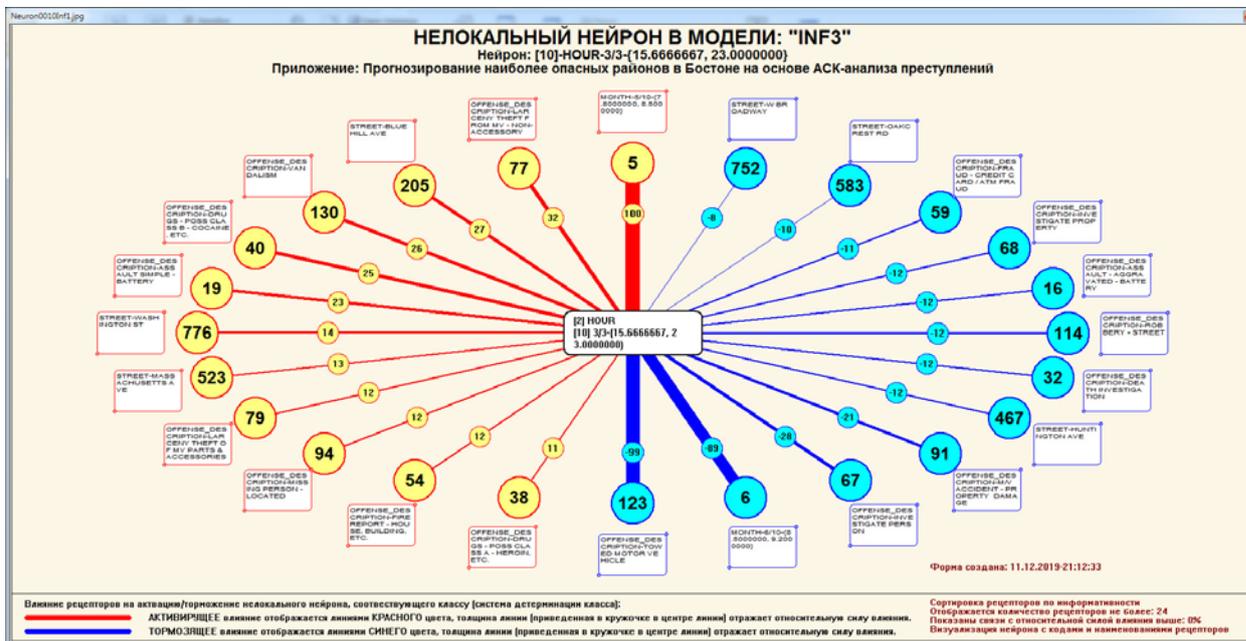


Рисунок 25 – Графическое отображение нелокального нейрона в модели INF3

Дополнение модели нейрона связями факторов позволяет построить классическую когнитивную карту ситуации (будущего состояния АОУ). Детальная внутренняя структура любой связи отображается в форме инвертированной когнитивной диаграммы (Рисунок 26). Необходимо отметить, что все указанные графические формы генерируются системой «Эйдос» автоматически в соответствии с созданной моделью [10].

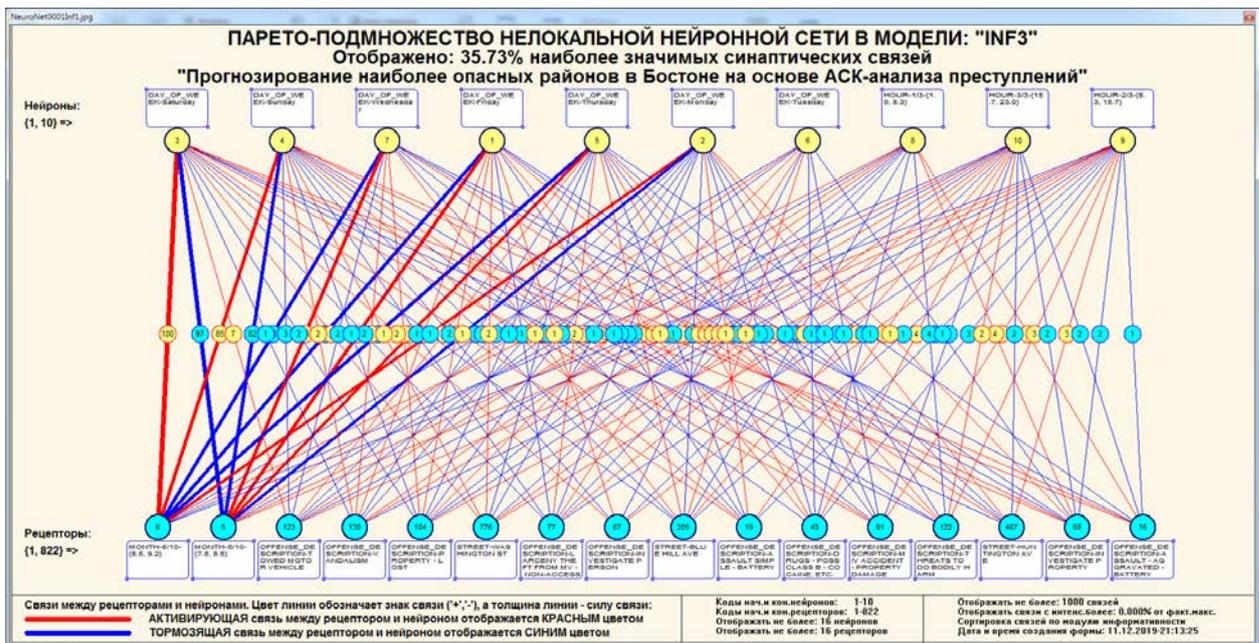


Рисунок 26 – Парето-подмножество нелокальной нейронной сети в модели INF3

Модель знаний системы «Эйдос» относится к нечетким декларативным гибридным моделям и объединяет в себе некоторые особенности нейросетевой и фреймовой моделей представления знаний. Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

От фреймовой модели представления знаний модель системы «Эйдос» отличается своей эффективной и простой программной реализацией, полученной за счет того, что разные фреймы отличаются друг от друга не набором слотов и шпаций, а лишь информацией в них. Поэтому в системе «Эйдос» при увеличении числа фреймов само количество баз данных не увеличивается, а увеличивается лишь их размерность.

На рисунке (Рисунок 27) приведены когнитивная диаграмма и агломеративная дендрограмма когнитивной кластеризации значений факторов и график изменения межкластерных расстояний, полученные на основе той же матрицы сходства признаков по их смыслу, что и в когнитивных диаграммах.

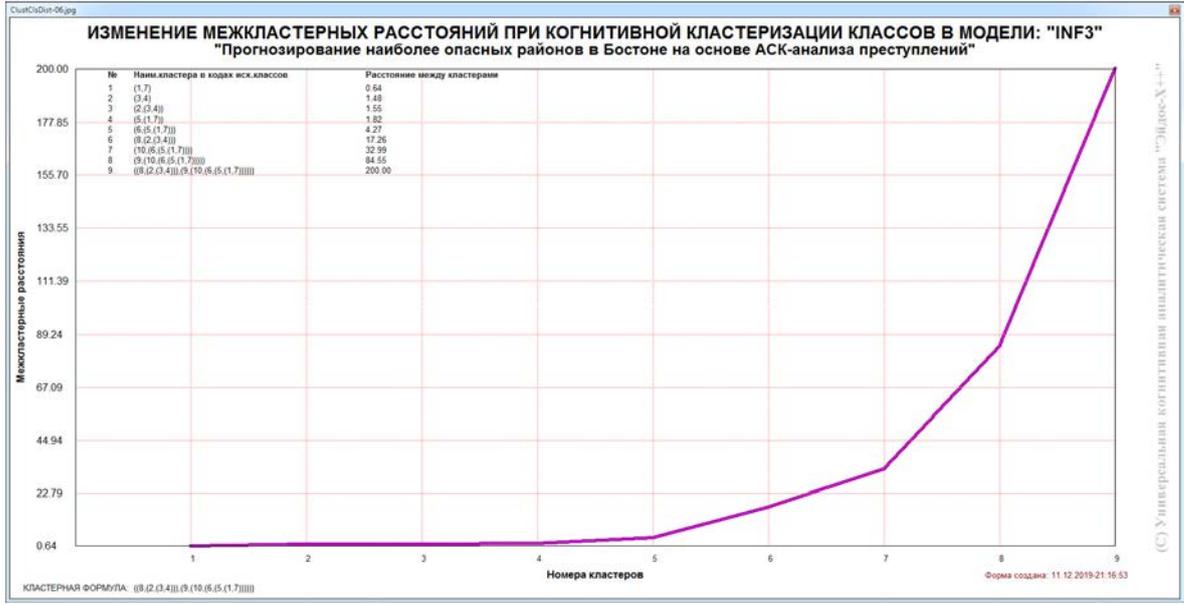
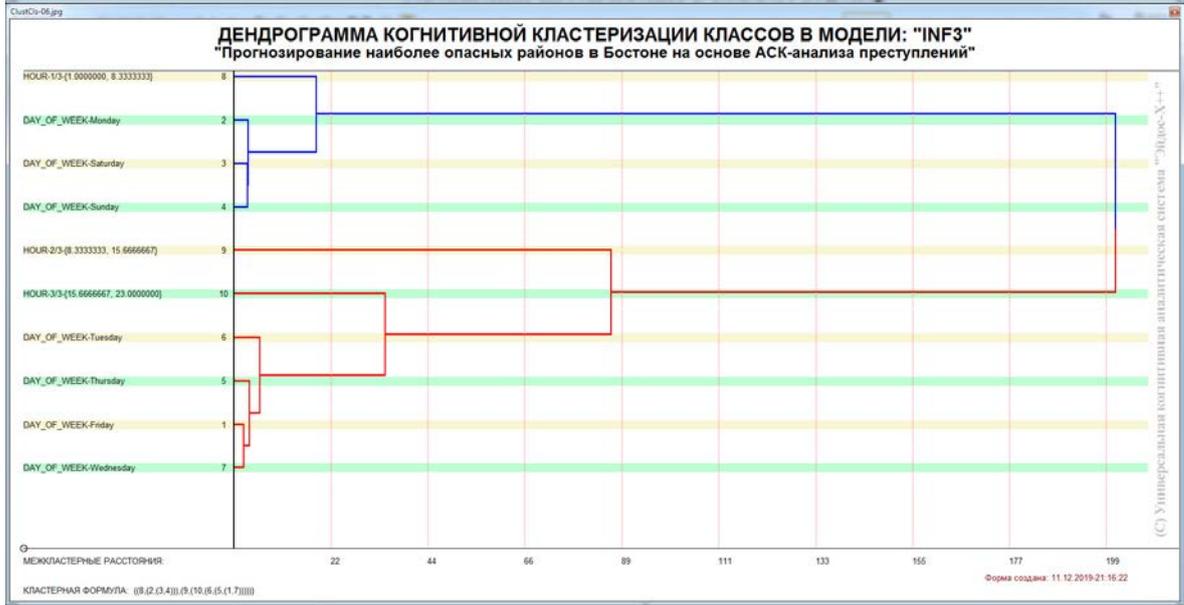
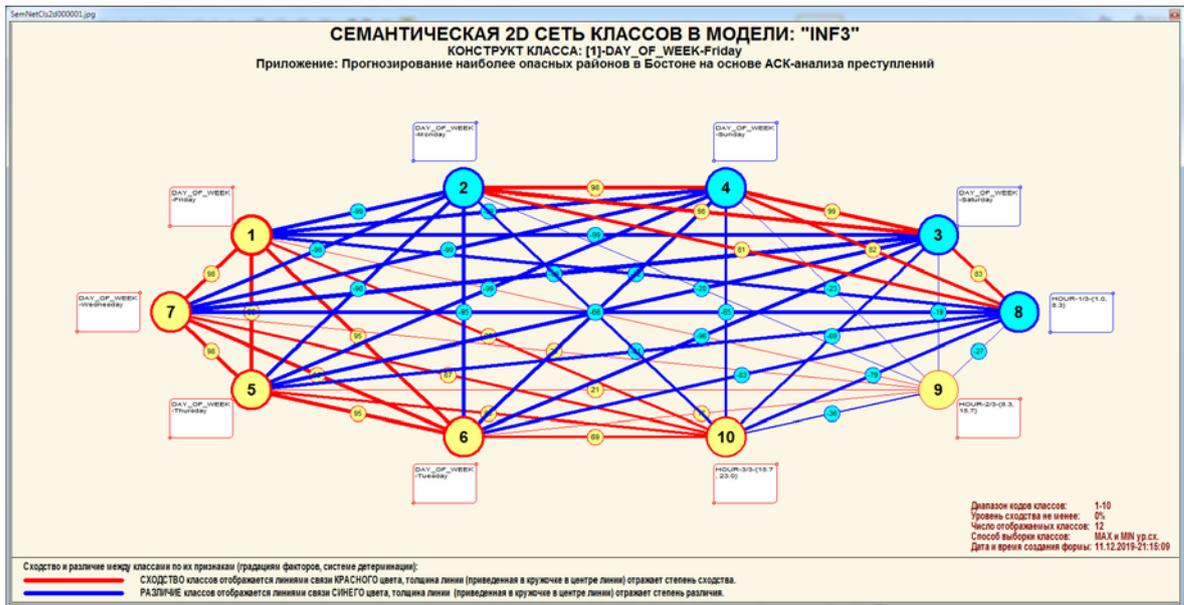


Рисунок 27 – Когнитивная диаграмма и агломеративная дендрограмма когнитивной кластеризации значений факторов и график изменения межкластерных расстояний

## **2.5. SWOT и PEST матрицы и диаграммы**

При принятии решений определяется сила и направление влияния значений факторов на принадлежность состояний объекта моделирования к тем или иным классам, соответствующим различным будущим состояниям. По сути, это решение задачи SWOT-анализа [11].

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.

Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных значений параметров на присутствие человека в помещении.

В системе «Эйдос» в режиме 4.4.8 поддерживается решение этой задачи. При этом выявляется система детерминации заданного класса, т.е. система значений факторов, обуславливающих переход объекта моделирования и управления в состояние, соответствующее данному классу, а также препятствующих этому переходу.

На рисунках (Рисунок 28) приведены SWOT-диаграммы, наглядно отражающие силу и направление влияния различных значений параметров на присутствие человека в помещении.

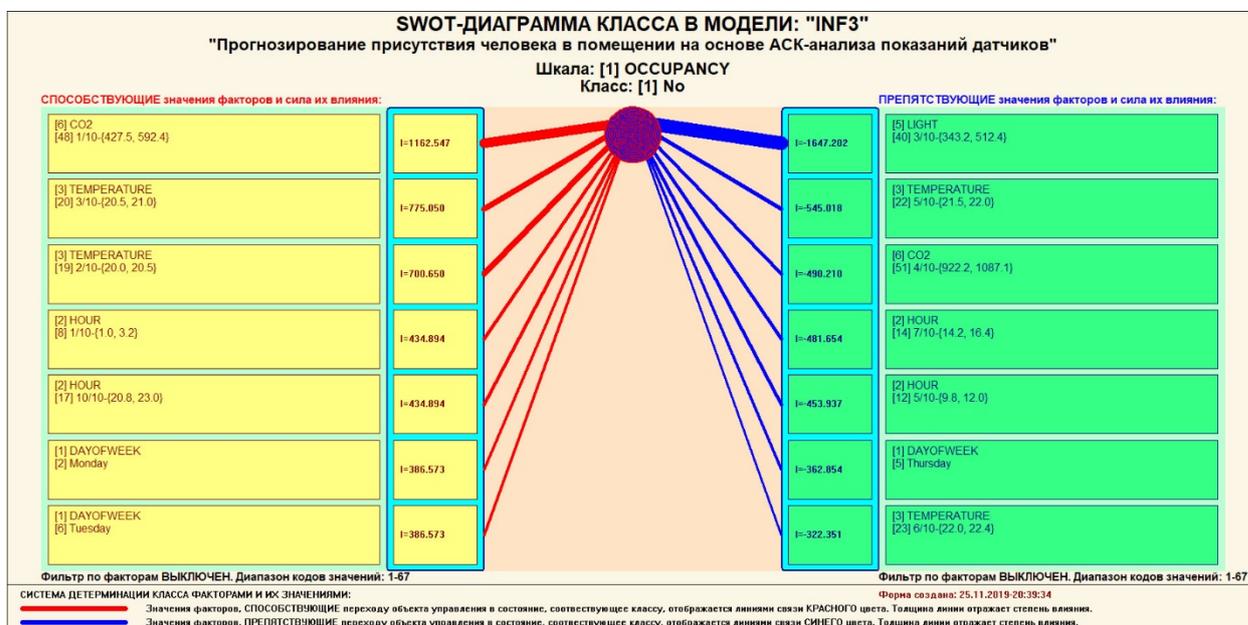


Рисунок 28 – SWOT-диаграмма класса DAY\_OF\_WEEK в модели INF3

## Заключение

Как показывает анализ результатов численного эксперимента предложенное и реализованное в системе «Эйдос» решение поставленных задач является вполне эффективным, что позволяет обоснованно утверждать, что цель работы достигнута, поставленная проблема решена.

Так как существует множество систем искусственного интеллекта, то 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 очень удобно использовать общедоступную базу репозитория KAGGLE.

В данной курсовой работе приводится развернутый пример использования базы данных репозитория KAGGLE для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом

наиболее достоверной в данном приложении оказались модель INF3, основанная на частном критерии Хи-квадрат, разности между фактическими и ожидаемыми абсолютными частотами при интегральном критерии «Сумма знаний». Точность модели составляет 0,915, что является отличным результатом. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием. Также необходимо обратить внимание, что статистические модели в данном приложении дают примерно на 21% более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, что, как правило, наблюдается и в других приложениях. Этим и оправдано применение моделей знаний.

В результате проделанной работы, с помощью системы «Эйдос» были созданы статистические и системно-когнитивные модели, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы классов по различным значениям показателей присутствия человека в помещении, изучено влияние значений различных параметров на эти прогнозы, и, на основе этого, решены задачи квалиметрии, классификации и исследования моделируемой предметной области путем исследования ее модели.

Данная работа является примером успешного применения когнитивных и информационных технологий для решения задач прогнозирования наиболее опасных районов на основе АСК-анализа базы преступлений в Бостоне.

## Список литературы

1. Occupancy Detection Data Set // [Электронный ресурс] intuit.ru. URL: <https://archive.ics.Kaggle.edu/ml/datasets/Occupancy+Detection+#>
2. Возможность использования систем автоматического управления освещением в длинном коридоре: XI межд. науч.-практ. конф. (Барнаул, 04-05.02.2016 г.) / Баранова И.А., Кондратьева Н.П., Большин Р.Г., Краснолуцкая М.Г., Коростелёв Д.В. Барнаул: АГАУ, 2016. С. 10-12.
3. Основы построения системы «Умный дом» // [Электронный ресурс] intuit.ru. URL: <https://www.intuit.ru/studies/courses/644/500/lecture/11366>
4. Сайт профессора Е.В. Луценко // [Электронный ресурс] intuit.ru. URL: <http://lc.kubagro.ru/>
5. Луценко Е.В., Боровко А.Ю. ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА И КЛАССОВ СОЛНЕЧНЫХ ВСПЫШЕК НА ОСНОВЕ ИХ ПРЕДЫСТОРИИ ПО ДАННЫМ РЕПОЗИТОРИЯ KAGGLE С ПРИМЕНЕНИЕМ АСК-АНАЛИЗА И ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ «ЭЙДОС» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/99.pdf>
6. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.
7. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена

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.п.л.

8. 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.

9. Луценко Е.В. Метод визуализации когнитивных функций - новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2011. - №03(67). С. 240 - 282. - Шифр Информрегистра: 0421100012\0077, 2,688 у.п.л. - Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>

10. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.

11. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. –

№07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

12. E. Hailemariam. Real-time occupancy detection using decision trees with multiple sensor types / Ebenezer Hailemariam, Rhys Goldstein, Ramtin Attar, Azam Khan // SimAUD '11 Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design, Pages 141-148. Boston, Massachusetts – April 03 - 07, 2011