

7. **Светлов В.А.** Философия и методология науки: учеб. пособие. Ч.1 / В.А. Светлов, И.А. Пфаненштиль. - Красноярск: Сибирский федеральный ун-т, 2011. – С. 356.
8. **Гублер Е.В.** Вычислительные методы анализа и распознавания патологических процессов / Е.В. Гублер. — Л.: Медицина, 1978. — 296 с.
9. **Вальд А.** Последовательный анализ: пер. с англ.; под ред. Б.А. Севастьянова.- Москва: Физматгиз, 1960. – 328 с.
10. **Кульбак С.** Теория информации и статистика: пер. с англ.; под ред. А.Н. Колмогорова. - Москва: Наука, 1967. – 408 с.

ГИПОТЕТИКО-ИНДУКТИВНЫЙ МЕТОД КАК ОСНОВА ИНФОРМАЦИОННО-ОБУЧАЮЩЕЙ ЭКСПЕРТНОЙ СИСТЕМЫ В СЛУЧАЯХ СМЕРТИ ОТ РАНЕНИЙ ОСТРЫМИ ПРЕДМЕТАМИ

Зосименко В.В., Зольникова А.Ю.

Резюме. В статье изложены подходы к построению информационно-учебной экспертной системы классификации преступника, как инструмента для анализа случаев убийств с использованием острых предметов. Нами был установлен и проверен алгоритм последовательной диагностической процедуры, основанный на использовании формулы Байеса, как один из возможных среди технологий Data Mining. В дальнейшем, на основе гипотетико-индуктивного метода, указанный подход будет применен для построения информационно-учебной экспертной системы, заполняя которую известными элементами криминалистической характеристики преступления, будем получать характеристики (пол, возраст) преступника с определенной статистической достоверностью.

Ключевые слова: острые предметы, информационно-учебная экспертная система, статистический анализ, криминалистическая характеристика преступления, алгоритм последовательной диагностической процедуры.

HYPOTHETICAL-INDUCTIVE METHOD AS A BASIS INFORMATION AND TRAINING EXPERT SYSTEM IN CASES OF DEATH FROM INJURY WITH A SHARP OBJECT

Zosimenko V.V., Zolnikova A.Yu.

Resume. The article describes the approaches to the construction of educational information and expert system classification of the offender, as a tool for the analysis of murders with sharp objects. We have installed and tested an algorithm sequential diagnostic procedure based on the use of Bayes' formula, as one of the possible technologies including Data Mining. In the future it will be used for the construction of information and training expert system, which is known for filling elements of criminalistics characterization of crime, will be obtained characteristics (sex, age) of the criminal with a certain statistical confidence. This information and training expert system will be developed in the PHP programming language with the user interface, and will be used in the educational process at the department of forensic medicine, and in the future and make a difference in the practical work of the forensic experts, especially during examinations in cases of violent death with a sharp object.

Keywords: sharp objects, information and training expert system, statistical analysis, criminalistics characterization of crime, algorithms of sequential diagnostic procedure.

УДК 340.64:343.97:616-035.8

МЕТОДИ ОБРОБКИ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ КЛАСУ ЗАЛЕЖНОЇ ЗМІННОЇ ПРИ АНАЛІЗІ ВИПАДКІВ ВБИВСТВ У М. КИЄВІ З ВИКОРИСТАННЯМ ГОСТРИХ ПРЕДМЕТІВ

©В. В. Зосіменко¹, А. Ю. Зольнікова²

¹Київське міське клінічне бюро судово-медичної експертизи

²Національна медична академія післядипломної освіти
імені П. Л. Шупика, Київ

Резюме. Створення експертної системи класифікації злочинця, як інструменту для аналізу випадків вбивств з використанням гострих предметів, залишається актуальною темою сьогодення. Протягом 2011-2014 років, нами було проведене ретроспективне дослідження 849 неселективних випадків смерті у м. Києві від поранень, які спричинені гострим предметом. Зокрема, було наведено обґрунтування того, що на сьогоднішній день для побудови інформаційно-навчальної експертної

системи (ИНЕКС) з метою аналізу подібного роду даних, доцільно використовувати технології Data Mining (видобуток даних, вилучення інформації, інтелектуальний аналіз даних). Однак, оптимального використання алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств у м. Києві з використанням гострих предметів, до сих пір не започатковано. Тому, з метою визначення оптимального використання алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств у м. Києві з використанням гострих предметів і було проведене наше дослідження. За результатами власних досліджень викладені підходи до вибору алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств з використанням гострих предметів. Нами був визначений оптимальний алгоритм такої статистичної обробки даних - використання технології Data Mining, що в подальшому буде застосоване для побудови системи, заповнюючи яку відомими елементами криміналістичної характеристики злочину, будемо отримувати характеристики злочинця з певною статистичною вірогідністю. Для пошуку закономірностей у великих обсягах неоднорідних та багатовимірних даних, які неочевидні (не виявляються стандартними методами обробки інформації або експертним шляхом), об'єктивні (повністю відповідають дійсності, на відміну від експертної саме думки, яке завжди є суб'єктивною) та корисні, на практиці доцільно використовувати технології Data Mining, які є оптимальні для з'ясування алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств у м. Києві з використанням гострих предметів.

Ключові слова: гострі предмети, статистичні методи, Data Mining, криміналістична характеристика злочину.

ВСТУП

Створення експертної системи класифікації злочинця, як інструменту для аналізу випадків вбивств з використанням гострих предметів, залишається актуальною темою сьогодення. Протягом 2011-2014 років, нами було проведене ретроспективне дослідження 849 неселективних випадків смерті у м. Києві від поранень, які спричинені гострим предметом [1-3]. Зокрема, ми навели обґрунтування того, що на сьогоднішній день для побудови інформаційно-навчальної експертної системи (ИНЕКС) з метою аналізу подібного роду даних, доцільно використовувати технології Data Mining (видобуток даних, вилучення інформації, інтелектуальний аналіз даних). Однак, оптимального використання алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств у м. Києві з використанням гострих предметів, до сих пір не започатковано.

Метою даної роботи було визначення оптимального використання алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств у м. Києві з використанням гострих предметів.

Матеріали та методи дослідження. Об'єктом дослідження були 1068 «Актів судово-медичного дослідження трупів» та «Висновків експерта» з приводу вбивств із використанням гострого предмету, отриманих з архіву Київського міського бюро судово-медичної експертизи. Глибина дослідження складала 15 років (1997-2011 р.р.). Серед них - 930 випадків з наявністю даних про злочинця (вік, стать), з яких нами були обрані 849 випадків, у яких одному злочинцю відповідала одна жертва (труп). Отримані результати були оброблені стандартними методами варіаційної статистики.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ ТА ЇХ ОБГОВОРЕННЯ

Дані великого об'єму і великої розмірності представляються позбавленими структури та зв'язків, і для їх обробки використовуються технології *Data Mining*, в основі яких є концепція шаблонів (patterns), що представляють собою закономірності, які притаманні підвбіркам даних, та є потужним апаратом сучасної аналітики для виявлення прихованих закономірностей і побудови передбачуваних моделей, які ґрунтуються не на умоглядних міркуваннях, а на реальних даних [4].

Термін *Data Mining* отримав свою назву з двох понять: пошуку цінної інформації у великій базі даних (data) і видобування гірської руди (mining). Обидва процеси вимагають або просіювання величезної кількості сирого матеріалу, або розумного дослідження і пошуку потрібних цінностей.

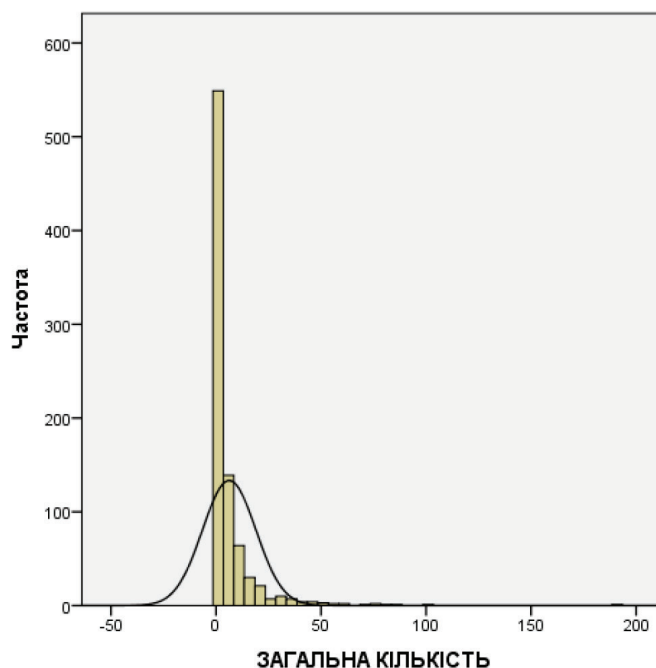
Термін *Data Mining* часто перекладається як видобуток даних, вилучення інформації, інтелектуальний аналіз даних. Традиційні методи аналізу даних: статистичні методи і розвідувальний аналіз, що становить основу оперативної аналітичної обробки даних - *OnLine Analytical Processing (OLAP)*, в основному використовуються для перевірки заздалегідь сформульованих гіпотез (*verification-driven data mining*), в той час як одне з основних положень *Data Mining* – це пошук неочевидних закономірностей.

Також необхідно враховувати і те, що для виявлення взаємозв'язків в даних, більшість статистичних методів використовують концепцію усереднення за вибіркою, яка приводить до операцій над неіснуючими величинами, тоді як *Data Mining* оперує реальними значеннями. Варто відзначити, що такий критерій, як критерій Стьюдента, який найбільш часто зустрічається при порівнянні вибірок, заснований на математичних допущеннях про те що, порівнювані вибірки повинні відбуватися з нормально розподілених сукупностей, і дисперсії порівнюваних генеральних сукупностей повинні бути рівні. У реальній практиці ці дві умови рідко виконуються.

Як приклад, наведемо такі наші дані, як те, що у випадках смерті від поранень, які спричинені гострим предметом у м. Києві (N=849), найбільш часто у 362 (42,6%) випадків смерть настала в результаті завдання 1 поранення (Me(медіана)=2; Mo(мода)=1; Q1(перший квартиль)=1; Q3(третій квартиль)=6), тоді як середнє значення

складає $M=6,3$ при стандартному відхиленні $\sigma=12,7$, що означає явну асиметрію розподілу показника у виборці (Мал. 1).

Data Mining - це не один, а сукупність різних методів виявлення знань, які пов'язані із аналізом великих об'ємів даних. Вибір залежить від типу наявних даних і від того, яку інформацію необхідно отримати.



Мал. 1. Гістограма розподілу випадків вбивств за кількістю нанесених поранень на тлі кривої нормального (гаусового) розподілу

До методів і алгоритмів *Data Mining* відносять наступне: штучні нейронні мережі; дерева рішень; символічні правила; методи найближчого сусіда і k -найближчого сусіда; метод опорних векторів; байєсовські мережі; лінійна регресія; кореляційно-регресійний аналіз; ієрархічні методи кластерного аналізу; неієрархічні методи кластерного аналізу, в тому числі алгоритми k -середніх і k -медіан; методи пошуку асоціативних правил, такі як алгоритм *Apriori*; метод обмеженого перебору, еволюційне програмування і генетичні алгоритми, різноманітні методи візуалізації даних та безліч інших.

Data Mining, як правило, складається з трьох стадій:

1. Виявлення закономірностей (вільний пошук). На стадії вільного пошуку здійснюється дослідження набору даних з метою пошуку прихованих закономірностей.

2. Використання закономірностей, які були виявлені, для передбачення невідомих значень (прогностичне моделювання). Прогностичне моделювання використовує результати роботи першої стадії. В процесі прогностичного моделювання вирішуються завдання класифікації і прогнозування. При вирішенні завдання класифікації результати роботи першої стадії (індукції правил) використовуються для віднесення нового об'єкту, з певною впевненістю, до одного з відомих, визначених класів на підставі відомих значень. *Data Mining*-средства можуть видавати велику кількість результатів, які статистично недостовірні. Щоб цього уникнути, необхідна перевірка адекватності отриманих моделей на тестових даних, і на додаток до першої стадії (вільного пошуку), вводять «підстадію» валідації. Мета валідації - перевірка достовірності знайдених закономірностей.

3. Аналіз винятків – стадія, яка призначена для виявлення та пояснення аномалій, які знайдені в закономірностях - виявлення відхилень (deviation detection). Для виявлення відхилень необхідно визначити норму, яка розраховується на стадії вільного пошуку.

Розглянемо задачу визначення передбачуваних статі та віку злочинця за наявними характеристикам злочину - це завдання класифікації. Мета процесу класифікації в нашому масиві випадків - побудувати модель, яка використовує атрибути, що прогнозують (відомі предиктори), в якості входних параметрів та отримує значення залежного атрибуту (стать та вік злочинця, які прогнозуються).

У разі класифікації припускається клас залежної змінної, і для цього використовують методи, які розкладають множини об'єктів на класи за певним критерієм. Основні з них:

- класифікація за допомогою дерев рішень;
- байєсовська (наївна) класифікація;
- класифікація за допомогою штучних нейронних мереж;

- класифікація методом опорних векторів;
- класифікація за допомогою метода найближчого сусіда.

Сьогодні існує великий вибір програмних засобів як комерційних, так і с GNU (General Public License) ліцензією, які дозволяють використовувати будь-який з вище перелічених методів. Для рішення подібних завдань також можна використовувати різні обчислювальні методи розпізнавання. Розглянемо один з таких методів - так звану послідовну діагностичну процедуру розпізнавання, в основі якої лежить метод секвенціального (послідовного) аналізу, який був розроблений А. Вальдом та набув поширення в низці областей техніки, а також біології і медицини [5, 6].

Обґрунтування використання послідовного статистичного аналізу для діагностики, і відповідні методичні прийоми його використання з цією метою були дані А.А.Генкіним и Е.В.Гублером [7]. Алгоритм послідовної діагностичної процедури впливає з основних теорем теорії ймовірностей і, зокрема, з формули Байєса, яка дуже часто використовується при обчислювальній діагностиці [8].

При введенні певних обмежень діагностичної задачі і початкових умов, формула Байєса приймає вигляд:

$$\text{поріг В} < \frac{P(x_1^1/A)}{P(x_1^1/B)} \times \frac{P(x_2^3/A)}{P(x_2^3/B)} \cdots \frac{P(x_j^i/A)}{P(x_j^i/B)} < \text{поріг А},$$

Де $\frac{P(x_j^i/A)}{P(x_j^i/B)}$ - відношення правдоподібності для виявленої i -й градації j -ої ознаки при належності до групи А, до ймовірності цієї ж градації тієї ж ознаки при належності до групи В. Таким чином, збір діагностичної інформації та перемноження відносин ймовірностей знайдених ознак продовжують до тих пір, поки правильна нерівність, тобто поки не

досягнутий «поріг А» або «поріг В».

У загальному випадку величини порогів для прийняття рішення з необхідним рівнем надійності за Вальдом визначають за формулою:

$$\text{поріг А} = \frac{1 - \alpha}{\beta}; \quad \text{поріг В} = \frac{\alpha}{1 - \beta},$$

де α и β – помилки першого та другого роду відповідно.

Таким чином, при послідовній процедурі розпізнавання для вибору гіпотези ми використовуємо тільки ту частину діагностичної інформації (у вигляді множення відносин правдоподібності), яка необхідна для досягнення певного порогу. Для зручності обчислень, множення відносин правдоподібності замінюють відповідним йому складанням логарифмів цих величин. Щоб ці логарифми представляли собою цілі числа, їх множать на 10 і округлюють з точністю до одиниці. Величину, яку при цьому отримують, називають діагностичним коефіцієнтом. Діагностичний коефіцієнт i градації ознаки x_j дорівнює:

$$\text{ДК}(x_j^i) = 10 \lg \frac{P(x_j^i/A)}{P(x_j^i/B)}.$$

Оскільки доданки використовуваної нерівності представляють собою діагностичні коефіцієнти відповідних діапазонів відповідних ознак, запишемо його у вигляді:

$$10 \lg \frac{\alpha}{1 - \beta} < \text{ДК}(x_1) + \text{ДК}(x_{21}) + \cdots + \text{ДК}(x_j) < 10 \lg \frac{1 - \alpha}{\beta}.$$

Діагностичні коефіцієнти вчислюють для усіх діапазонів ознак, при цьому записують їх у вигляді діагностичної таблиці строго в порядку інформативності, яка зменшується. Під диференціальною інформативністю ознаки розуміють ступінь відмінностей його розподілів при диференційованих належностях до груп А і В.

Чим сильніше розрізняються ці розподіли, тим більш інформації несе розглянута ознака. Якщо її розподіли при належності до груп А і В взагалі не перетинаються, інформативність ознаки нескінченно велика; ознака у всіх випадках дозволяє однозначно визначити, відноситься даний об'єкт до класу А або В. При використанні послідовної діагностичної процедури зручною мірою для оцінки інформативності ознак є міра Кульбака [9].

На відміну від критерію χ^2 та інших критеріїв статистичної значимості відмінностей, міра Кульбака дозволяє оцінити не вірогідність (достовірність) відмінностей між розподілами, а ступінь цих відмінностей.

Спочатку вчислюють інформативність градацій діапазонів ознаки x . Відповідно до формули Кульбака, величина інформативності I -ого діапазону i ознаки j дорівнює:

$$I(x_j^i) = ДК(x_j^i) \frac{1}{2} [P(x_j^i/A) - P(x_j^i/B)].$$

Однак, для визначення порядку використання ознак в діагностичній таблиці, необхідно обчислити інформативність не одного діапазону ознаки, а всієї ознаки. Інформативність всієї ознаки x_j дорівнює сумі інформативності її діапазонів. При складанні діагностичної таблиці, перше завдання, яке доводиться вирішувати, це завдання відбору найбільш інформативних ознак. Попередній відбір інформативних ознак для подальшої роботи можна проводити за допомогою будь-якого критерію відмінностей. У нашому випадку ми, в т.ч. посиляючись на алгоритм вибору одномірного статистичного аналізу [10], використовували критерій χ^2 та коефіцієнт чотирьох-польової кореляції ϕ . Ознаки, відмінності між якими в групах А і В не є статистично значущими, доцільно вилучити з подальшої роботи. Однак якщо інформативних ознак мало, виявляється доцільним залишити для подальшої роботи і ті ознаки, у яких відмінності між групами А і В недостатньо значущі. Остаточний висновок про інформативність всього набору ознак можна зробити тільки після перевірки складеної діагностичної таблиці на досить великій перевірочній групі спостережень. Якщо число помилок виявиться велике, можна надалі виключити найменш інформативні ознаки.

Як було зазначено у вступі цієї статті, нами були оцінені предиктори за ретроспективною базою неселективних випадків смерті у м. Києві від поранень, які спричинені гострим предметом. Є дані про інцидент: вік, стать злочинця і жертви, кількість ран і їх локалізація, місце скоєння злочину, характер дня та інші характеристики. Для кожного предиктора, з урахуванням частоти зустрічаємості у групах, були розраховані діагностичні коефіцієнти та їх інформативність. Виявилось, що з усіх вивчених параметрів, діагностичною інформативністю (більш 0,5) володіє група предикторів, які ми в подальшому і використовували для складання диференційно-діагностичної таблиці співставлення об'єкту з тим чи іншим класом.

ВИСНОВОК

Отже, існує великий вибір методів та програмних засобів їх реалізації для обробки великих обсягів даних, починаючи від давно відомих та використовуваних методів, які засновані на формулі Байєса, до сучасних генетичних алгоритмів. Вибір застосування конкретного методу залишається за дослідником. Для пошуку закономірностей у великих обсягах неоднорідних та багатовимірних даних, які неочевидні (не виявляються стандартними методами обробки інформації або експертним шляхом), об'єктивні (повністю відповідають дійсності, на відміну від експертної саме думки, яке завжди є суб'єктивною) та корисні, на практиці доцільно використовувати технології Data Mining, які є оптимальні для з'ясування алгоритмів статистичної обробки даних, як інструменту для обґрунтування результатів аналізу випадків вбивств у м. Києві з використанням гострих предметів.

Література

1. **Зосіменко В.В.** Аналіз випадків вбивств з використанням гострих предметів у м. Києві як основи для обґрунтування криміналістичної характеристики злочину/В.В. Зосіменко//Судово-медична експертиза. – 2011. - №3. – С. 16-20.
2. **Зосіменко В.В.** Статистичне моделювання ознак, які індивідуалізують особу за наслідками її агресивних дій з використанням гострих предметів /В.В. Зосіменко//Судово-медична експертиза. – 2014. - №2. – С. 23-27.
3. **Зосіменко В.В.** Експертний аналіз випадків вбивств із застосуванням гострих предметів як основа криміналістичної характеристики злочину / В.Д. Мішалов, В.В. Зосіменко //Інтегративна антропология. – 2014. - № 2(24). – С. 33-37.
4. **Дюк В.А.** Data Mining: учебный курс. / В.А. Дюк, А.П. Самойленко. – СПб: Питер, 2001. - 368 с.
5. **Башаринов А.Е.** Методы статистического последовательного анализа и их радиотехнические приложения. / А.Е. Башаринов, Б.С. Флейшман. – Москва: Советское радио, 1962. - 352 с.
6. **Урбах В.Ю.** Биометрические методы. / В.Ю. Урбах. – Москва: Наука, 1964. - 415 с.
7. **Гублер Е.В.** Применение непараметрических критериев статистики в медико-биологических исследованиях. / Е.В. Гублер, А.А. Генкин. – Ленинград: Медицина, 1973. - 141 с.
8. **Парин В.В.** Введение в медицинскую кибернетику / В.В. Парин, Р.М. Баевский. – Москва: Медицина, 1966. - 300 с.
9. **Кульбак С.** Теория информации и статистика: пер. с англ.; под ред. А.Н.Колмогорова. - Москва: Наука, 1967. – 408 с.
10. **Румянцев П.О.** Статистические методы анализа в клинической практике. Часть I. Одномерный статистический анализ / П.О. Румянцев, В.А. Саенко, У.В. Румянцева // Проблемы эндокринологии. - 2009. - № 55(5). - С. 48-55.

МЕТОДЫ ОБРАБОТКИ ДАННЫХ ДЛЯ ОПРЕДЕЛЕНИЯ КЛАССА ЗАВИСИМОЙ ПЕРЕМЕННОЙ ПРИ АНАЛИЗЕ СЛУЧАЕВ УБИЙСТВ В Г. КИЕВЕ С ИСПОЛЬЗОВАНИЕМ ОСТРЫХ ПРЕДМЕТОВ

В. В. Зосименко, А. Ю. Зольникова

Резюме. Создание экспертной системы классификации преступника, как инструмента для анализа случаев убийств с использованием острых предметов, остается актуальной темой сегодняшнего дня. В течение 2011-2014 годов, нами было проведено ретроспективное исследование 849 неселективных случаев смерти в г. Киеве от ранений, вызванных острым предметом. В частности, было приведено обоснование того, что для построения информационно-учебной экспертной системы (ИУЭС) с целью анализа подобного рода данных, целесообразно использовать технологии Data Mining (получение данных, извлечение информации, интеллектуальный анализ данных). Однако, оптимального использования алгоритмов статистической обработки данных, как инструмента для обоснования результатов анализа случаев убийств в г. Киеве с использованием острых предметов, до сих пор не начато. Цель - определение оптимального использования алгоритмов статистической обработки данных, как инструмента для обоснования результатов анализа случаев убийств в г. Киеве с использованием острых предметов. Материал и методы - архивные материалы Киевского городского клинического бюро судебно-медицинской экспертизы. Стандартные методы вариационной статистики. Результаты и выводы. В судебной медицине известен большой выбор методов и программных средств их реализации для обработки больших объемов данных, начиная от давно известных и используемых методов, основанных на формуле Байеса, в современных генетических алгоритмов. Выбор применения конкретного метода остается за исследователем. По результатам собственных исследований, изложены подходы к выбору алгоритмов статистической обработки данных, как инструмента для обоснования результатов анализа случаев убийств с использованием острых предметов. Нами был определен оптимальный алгоритм такой статистической обработки данных - использование технологии Data Mining, что в дальнейшем будет применено для построения системы, заполняя которую известными элементами криминалистической характеристики преступления, будем получать характеристики преступника с определенной статистической достоверностью. Для поиска закономерностей в больших объемах неоднородных и многомерных данных, которые неочевидны (не обнаруживаются стандартными методами обработки информации или экспертным путем), объективные (полностью соответствуют действительности, в отличие от экспертной которая чаще всего, субъективна) и полезные, на практике целесообразно использовать технологии Data Mining, которые являются оптимальными для выяснения алгоритмов статистической обработки данных, как инструмента для обоснования результатов анализа случаев убийств в г. Киеве с использованием острых предметов.

Ключевые слова: острые предметы, статистические методы, Data Mining, криминалистическая характеристика преступления.

METHODS OF DATA TO DETERMINE THE CLASS OF THE DEPENDENT VARIABLE IN THE ANALYSIS OF MURDERS IN KIEV WITH SHARP OBJECTS

Zosimenko V.V., Zolnikova A.Yu.

Summary. During 2011-2014 years we conducted a retrospective study of 849 non-selective deaths in Kiev from wounds, that were caused by a sharp object. The aim of this study was to determine the optimal algorithms of the statistical processing of a data, as a tool to justify the results of the analysis of cases of murders using sharp objects in Kiev. **Materials and methods.** Generally, the object of the study were 1068 "Acts of the forensic investigation of corpses" and "expert opinion" about the killings with using a sharp object, obtained from the archives of the Kiev City Bureau of forensics. Depth study was 15 years (1997-2011 years). Among them - 930 cases of the availability of data on the offender (age, gender), of which we selected 849 cases, in which one offender met one victim (corpse).

Results. The article presents selected our approaches to the definition of statistical data processing algorithms as a tool to support the analysis of murders with sharp objects.

Conclusion. We have determined the optimal algorithm such statistical data - the use of technology Data Mining, which will continue to be applied for the construction of the system, which is known for filling elements of criminalistics characteristics of crime, the characteristics of the offender will get a certain statistical confidence.

Key words: Sharp, Statistical Methods, Data Mining, Criminalistics Characteristics of Crime.