УДК 004.942

РЕШЕНИЕ ЗАДАЧИ КЛАССИФИКАЦИИ СОТРУДНИКОВ ПО УРОВНЮ ДОСТУПА К КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ НА ЗАКРЫТОМ ПРЕДПРИЯТИИ МЕТОДОМ ДИСКРИМИНАНТНОГО АНАЛИЗА

Мироненко Владимир Андреевич, студент, направление подготовки 09.03.01 Информатика и вычислительная техника, Оренбургский государственный университет, Оренбург
e-mail: v.miron2002@yandex.ru

Научный руководитель: **Костин Владимир Николаевич**, доктор технических наук, доцент, доцент кафедры программного обеспечения вычислительной техники и автоматизированных систем, Оренбургский государственный университет, Оренбург
e-mail: vladimirkostin5@mail.ru

Аннотация. *В современном мире наблюдается рост утечек различного рода информации. Так, с 2022 по 2023 годы многократно увеличился поток утечек данных, содержащих коммерческую тайну, поэтому задачи, ставящие целью минимизировать данный процесс, являются актуальными. Целью исследования является создание математической модели, описывающей зависимость между должностью сотрудника и уровнем доступа к конфиденциальной информации, что необходимо для оценки степени принадлежности информации к определенному классу конфиденциальности и возможности мониторинга ее дальнейшего перемещения. Для решения поставленной задачи выбран дискриминантный анализ. Полученные результаты могут быть использованы для усиления мер безопасности и разработки новых процедур контроля доступа к конфиденциальной информации на закрытом предприятии.*

Ключевые слова: *дискриминантный анализ, информационная безопасность, утечка данных, аналитическое программное обеспечение, классификация данных.*

Для цитирования: Мироненко В. А. Решение задачи классификации сотрудников по уровню доступа к конфиденциальной информации на закрытом предприятии методом дискриминантного анализа // Шаг в науку. – 2024. – № 4. – С. 63–67.

SOLVING THE PROBLEM OF LEAKAGE OF CONFIDENTIAL INFORMATION IN A CLOSED ENTERPRISE BY THE METHOD OF DISCRIMINANT ANALYSIS

Mironenko Vladimir Andreevich, student, training program 09.03.01 Computer Science and Computer Engineering, Orenburg State University, Orenburg
e-mail: v.miron2002@yandex.ru

Research advisor: **Kostin Vladimir Nikolaevich**, Doctor of Technical Sciences, Associate Professor, Associate Professor of the Department of Computer Engineering and Automated Systems Software, Orenburg State University, Orenburg
e-mail: vladimirkostin5@mail.ru

Abstract. *In the modern world, there is an increase in leaks of various kinds of information. Thus, from 2022 to 2023, the flow of data leaks containing trade secrets increased almost twice, so the tasks aimed at minimizing this process are relevant. The aim of the study is to create a mathematical model describing the relationship between the position of an employee and the level of access to confidential information, which is necessary to assess the degree of belonging of information to a certain class of confidentiality and the possibility of monitoring its further movement. Discriminant analysis was chosen to solve the problem. The results obtained can be used to strengthen security measures and develop new procedures for controlling access to confidential information in a closed enterprise.*

Key words: *discriminant analysis, information security, data leak, analytical software, data classification*.

Cite as: Mironenko, V. A. (2024) [Solving the problem of leakage of confidential information in a closed enterprise by the method of discriminant analysis]. *Shag v nauku* [Step into science]. Vol. 4, pp. 63–67.

В современном мире информация является важнейшим стратегическим ресурсом. Согласно статистическим данным[1], с 2022 по 2023 годы общемировой рост утечек информации увеличился на 61,5%, «почти втрое возросла доля утечек сведений, составляющих коммерческую тайну (документы стратегического плана, секреты производства и т. д.). В 2023 году этот показатель составил 33,1%». В научных источниках [1–6; 8] рассматриваются различные вопросы защиты информации от несанкционированного доступа. Например, в работе [7] предлагается инженерное решение задачи организации хранения, передачи, поддержки поиска и воспроизводства научно-технической информации, а именно: использование цифровой платформы, ориентированной на контроль и управление динамическими информационными объектами с большой интеллектуальной и творческой составляющей, описываются основные элементы такой платформы. В исследовании [3] рассматривается задача выявления фактов утечки паролей на основе анализа динамики клавиатурного ввода. Авторы приходят к выводу, что «при многократном вводе фиксированных символьных последовательностей у пользователя формируется специфичная для него манера ввода, стабилизирующаяся с течением времени». На основе анализа научных работ авторы исследования [4] описывают основные угрозы информации и факты утечки конфиденциальной информации. Однако, несмотря на разносторонний подход к решению вопросов, связанных с защитой информации, задача предотвращения утечки данных остается актуальной.

Постановка задачи: классификация сотрудников закрытого предприятия по уровню доступа к конфиденциальной информации для оценки возможности мониторинга ее дальнейшего перемещения. Данная задача является первым этапом прогнозирования возможности утечки конфиденциальной информации на закрытом предприятии. Исходные данные представлены в виде информационной двумерной матрицы (таблица 1) [6]. Столбцы матрицы – зашифрованные типы информации. Рассматриваются следующие типы информации: $А_1$, $А_2$, $А_3$, $Б_1$, $Б_2$, $В_1$, $В_2$, $В_3$, $Г_0$, $Г_1$, $Г_2$, $Г_3$. Строки – носители этой информации. Номер строки – код должности, которой соответствует определенный уровень доступа к каждому типу информации. Поле матрицы – уровень владения информацией, определяется по шкале: 6 – самый высокий уровень владения информацией; 0 – самый низкий уровень, соответствует полному отсутствию доступа к данной информации.

Таблица 1. Исходные данные

Класс конфиденциальности	Код должности	Типы информации											
		$А_1$	$Б_1$	$А_2$	$Б_2$	$А_3$	$В_1$	$В_2$	$Г_3$	$Г_0$	$В_3$	$Г_1$	$Г_2$
0	1	3	2	1	1	3	6	6	2	1	2	1	1
0	2	3	2	1	1	3	6	6	1	1	2	1	1
0	3	1	2	1	2	2	2	2	1	1	1	0	1
0	4	1	2	0	0	1	1	1	1	1	1	0	1
0	5	1	1	2	1	1	1	1	1	1	1	1	1
0	6	2	2	1	2	2	2	1	1	0	1	1	1
0	7	1	1	1	1	2	1	1	1	2	1	1	1
0	8	2	1	2	1	3	2	1	1	2	2	2	1
1	9	1	1	1	1	2	1	1	6	6	5	6	6
1	10	1	1	1	1	1	3	1	6	4	3	3	3
1	11	1	1	1	2	1	1	1	3	4	3	6	6
1	12	1	1	1	1	1	1	1	3	3	3	3	6
1	13	0	1	0	0	0	0	0	2	3	2	6	4
1	14	1	1	0	0	2	1	0	3	3	6	3	4
2	15	6	3	6	2	6	2	1	2	1	2	1	1
2	16	3	6	2	6	6	2	1	2	1	2	1	1
2	17	5	2	6	2	5	2	1	1	1	2	1	1
2	18	3	6	1	5	4	4	3	1	1	2	1	1

Источник: разработано автором на основе работы [6]

[1] Утечки информации в мире, 2022–2023 годы // Экспертно-Аналитический центр InfoWatch. – URL: https://ict.moscow/research/utechki-informatsii-v-mire-2022-2023-gody/ (дата обращения: 01.04.2024).

Рисунок 1. Укрупненная схема алгоритма
Источник: разработано автором

Для решения поставленной задачи использовался дискриминантный анализ [9]. Его отличием от других методов многомерной классификации является то, что для существующих классов (множеств) формулируется правило, по которому в них распределяются новые единицы совокупности.

Входные данные представляют собой генеральную совокупность, состоящую из множества единиц наблюдения: кодов должностей сотрудников, обладающих различным уровнем доступа к конфиденциальной информации. Каждая единица наблюдения – сотрудник характеризуется несколькими признаками (дискриминантными переменными): типом информации, которой он владеет. Каждый из них есть x_{ij} – значение j-й переменной (j = 1, …, 12) для i-го объекта (i = 1, …, 18). Всё множество объектов разбито на 3 подмножества (класса) – уровней доступа сотрудника к конфиденциальной информации (таблица 1). Класс 0 соответствует низкому уровню доступа к конфиденциальной информации, класс 1 – среднему, класс 2 – высокому (таблица 1).

Основой дискриминантного анализа является построение «классифицирующей функции», которая максимизирует различия между классами, но минимизирует дисперсию внутри классов [9]. Она имеет следующий вид:

$$D_k = b_{k0} + b_{k1}X_1 + b_{k2}X_2 + \ldots + b_{kp}X_p, \qquad (1)$$

где

k – номер класса (для нашей задачи k = 1, …, 3),

b_{kj} – коэффициенты дискриминантной «классифицирующей» функции (для нашей задачи i = 1, …, 12),

p – количество дискриминантных переменных.

Коэффициенты для классифицирующих функций определяются с помощью таких вычислений [9]:

$$b_{ki} = (n_{*} - g)\sum_{j=1}^{p} a_{ij}X_{j k_{*}}, \qquad (2)$$

где

n_{*} – общее число наблюдений по всем классам,

g – число классов,

$X_{j k_{*}}$ – среднее значение переменной j в k-ом классе,

a_{ij} – элемент матрицы, обратной к внутригрупповой матрице сумм попарных произведений W [9].

$$W_{ij} = \sum_{k=1}^{g}\sum_{j=1}^{n_k}(X_{ikm} - X_{i k_{*}})(X_{jkm} - X_{j k_{*}}), \qquad (3)$$

где

n_k – число наблюдений в k-ом классе,

X_{ikm} – величина переменной i для m-го наблюдения в k-м классе.

Постоянный член (константа дискриминации) определяется так [9]:

$$b_{k0} = -0{,}5 \sum_{j=1}^{p} b_{kj} X_{j\,k}^{*} . \quad (4)$$

В ходе данного исследования была разработана программа, в которой реализованы следующие цели дискриминантного анализа:

– определение коэффициентов дискриминантных функций (по формулам (1)–(4));

– проверка существования между группами (классами) значимых различий с точки зрения независимых переменных;

– оценка точности классификации данных на группы.

В программе данные считываются из файла электронных таблиц MS Excel, результаты вычислений импортируются в тот же файл на отдельный лист. Укрупненная схема алгоритма представлена на рисунке 1.

Для каждого класса конфиденциальности получены следующие дискриминантные функции (индекс дискриминантной функции соответствует уровню конфиденциальности класса):

$$D_0 = 36{,}08 + 15{,}68 \cdot А_1 – 14{,}29 \cdot Б_1 – 11{,}26 \cdot А_2 + 11{,}16 \cdot Б_2 – 2{,}32 \cdot А_3 – 16{,}96 \cdot В_1 + \\ + 13{,}43 \cdot В_2 – 0{,}66 \cdot Г_3 + 4{,}39 \cdot Г_0 – 1{,}04 \cdot В_3 – 3{,}70 \cdot Г_1 – 12{,}30 \cdot Г_2$$

$$D_1 = -75{,}38 – 71{,}11 \cdot А_1 – 42{,}17 \cdot Б_1 – 18{,}64 \cdot А_2 – 32{,}69 \cdot Б_2 + 29{,}66 \cdot А_3 + 60{,}69 \cdot В_1 – 40{,}84 \cdot В_2 + \\ + 30{,}05 \cdot Г_3 – 51{,}08 \cdot Г_0 – 7{,}71 \cdot В_3 + 18{,}76 \cdot Г_1 + 59{,}68 \cdot Г_2$$

$$D_2 = -410.12 + 75{,}32 \cdot А_1 + 91{,}84 \cdot Б_1 + 50{,}48 \cdot А_2 + 26{,}71 \cdot Б_2 – 39{,}84 \cdot А_3 – 57{,}12 \cdot В_1 + \\ + 34{,}40 \cdot В_2 – 43{,}76 \cdot Г_3 + 67{,}85 \cdot Г_0 + 13{,}65 \cdot В_3 – 20{,}74 \cdot Г_1 – 64{,}93 \cdot Г_2$$

Геометрическая интерпретация результатов представлена на рисунке 2.

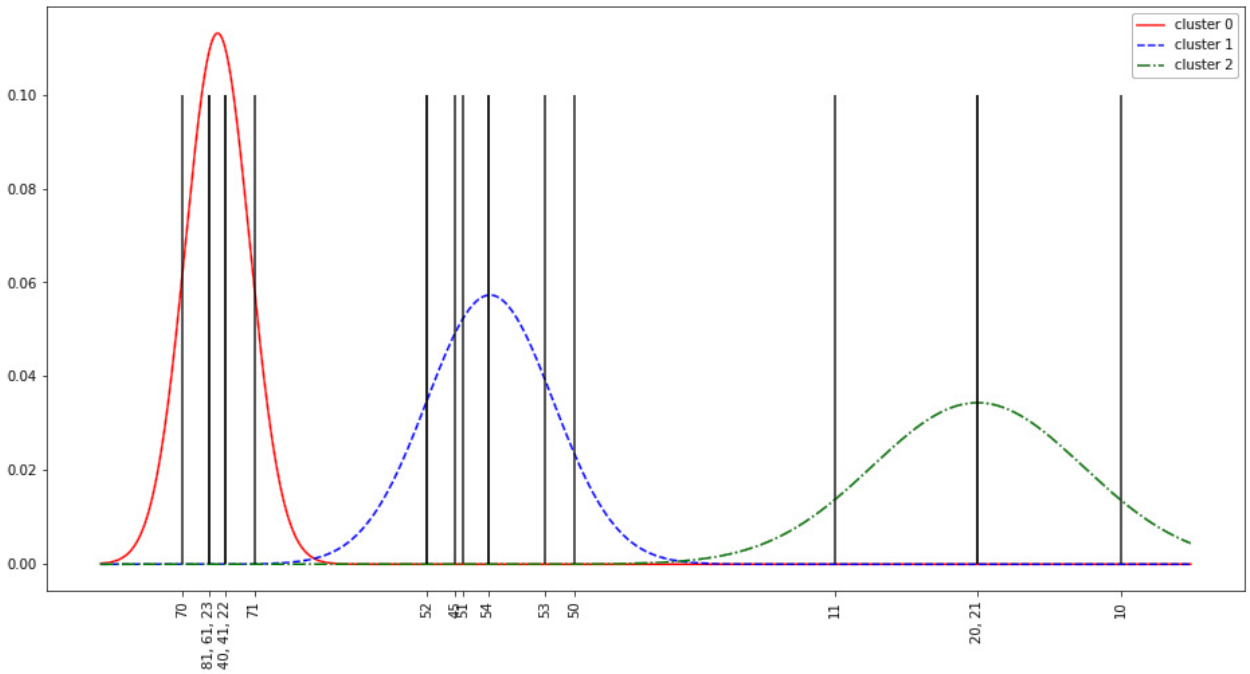

Рисунок 2. Геометрическая интерпретация результатов классификации сотрудников по уровню доступа к информации

Источник: разработано автором

Первый график (сплошной линией) отображает нулевой класс с низким уровнем конфиденциальности. Второй график (пунктирной линией) – первый класс среднего уровня конфиденциальности. Третий график (штрихпунктирной линией) соответствует второму классу конфиденциальности. Вертикальные сплош-

ные линии показывают конкретные наблюдения.

Полученные результаты исследования согласуются с результатами, полученными при использовании пакета прикладных программ Stastistica 10.0.

С помощью дискриминантной функции, используя функцию Лапласа для нормального распределения, можно определить положение информации в классе относительно математического ожидания и оценить вероятность принадлежности информации данному классу. Следующий этап исследования: спрогнозировать время перемещения информации за пределы класса. Данный момент интерпретируется, как наступление утечки информации, после которого необходимо выработать управляющие воздействия на перезагрузку модели безопасности предприятия – смена паролей, ключей, шифров, замков и т. д.

Программа, разработанная в рамках исследования, демонстрирует значительный потенциал для использования в различных областях, где требуется анализ больших объемов данных, решение задачи дискриминации (категорирования) и принятие на основе полученных результатов обоснованных решений. Результатом работы программы в рамках поставленной задачи являются дискриминантные функции для трех уровней конфиденциальности (низкого, среднего и высокого), на основе которых происходит классификация сотрудников по уровню доступа к различного рода информации. В дальнейшем, будет производиться оценка вероятности перехода сотрудника из одного класса конфиденциальности в другой, что позволит отследить вероятность утечки информации.

Литература

1. Булгаков Е. И., Зарудний А. В. Структура корпоративной системы предотвращения утечки конфиденциальной информации // Наукоемкие технологии и инновации (XXIII научные чтения): Международная научно-практическая конференция, посвященная 65-летию БГТУ им. В. Г. Шухова, Белгород, 29 апреля 2019 года. Том 9. – Белгород: Белгородский государственный технологический университет им. В. Г. Шухова, 2019. – С. 5–9.

2. Иняева К. Э., Сергеева И. А. Технические каналы утечки информации, причины утечки и способы защиты информации от утечки // Проблемы и перспективы развития российской экономики: сборник статей по материалам XII научно-практической конференции, Пенза, 19–20 декабря 2022 года. – М.: Издательство «Перо», 2023. – С. 94–96.

3. Карпычев В. Ю., Ляхманов Д. А., Охотников А. С. Выявление фактов утечки паролей методом анализа динамики клавиатурного ввода // Информационные системы и технологии ИСТ-2020: Сборник материалов XXVI Международной научно-технической конференции, Нижний Новгород, 24–28 апреля 2020 года. – Нижний Новгород: Нижегородский государственный технический университет им. Р. Е. Алексеева, 2020. – С. 559–565.

4. Коптева Л. Г., Рубан А. Г. Анализ способов защиты от утечек конфиденциальной информации // Наука и техника транспорта. – 2016. – № 4. – С. 86–90.

5. Костин В. Н., Боровский А. С. Метод оценки утечки конфиденциальной информации о функционировании системы защиты объекта информатизации по информационному критерию // Вестник компьютерных и информационных технологий. – 2016. – № 8(146). – С. 34–43. – https://doi.org/10.14489/vkit.2016.08.pp.034-043.

6. Костин В. Н., Даньшин Д. В. Метод оценки глубины прогноза развития (эволюции) характеристик сложных систем на основе энтропийного подхода // Информационные технологии. – 2015. – Т. 21, № 1. – С. 62–67.

7. Логинов Е. Л., Шкута А. А., Грабчак Е. П. Проблемы управления научно-технической информацией, ее защиты и поиска утечек данных // Экономика. Информатика. – 2021. – Т. 48, № 3. – С. 543–551. – https://doi.org/10.52575/2687-0932-2021-48-3-543-551.

8. Тарасова Т. М., Григанова Ю. С. Каналы утечки информации в коммерческом банке // Наука и образование транспорту. – 2019. – № 1. – С. 289–291.

9. Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др.; Под ред. И. С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.

Статья поступила в редакцию: 31.05.2024; принята в печать: 27.09.2024.

Автор прочитал и одобрил окончательный вариант рукописи.