

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВРЕМЕННОГО РЯДА КОЛИЧЕСТВА СКОМПРОМЕТИРОВАННЫХ ПЕРСОНАЛЬНЫХ ДАННЫХ В МИРЕ

Аннотация. В работе приведено описание статистических данных по количеству скомпрометированных персональных данных в период с 2013 по 2024 г. Показано, что с наибольшим коэффициентом корреляции данные описываются показательной функцией. Приведена попытка прогноза количества утечек информации на 2025 г.

Ключевые слова: математическое моделирование, прогноз, персональные данные, утечка.

Эпоха массовых кибер-атак, которая сегодня ведется стала тотальной во всем мире. Нет страны, в которой не фиксировались бы утечки информации национального масштаба. В 2023 г. экспертно-аналитическим центром было заявлено о 11549 утечек информации, произошедших в мире, что на 61,5 % больше, чем в 2022 г. [1]. Увеличение объема скомпрометированных персональных данных (ПДн) связано с повышением хакерских атак на фоне конфликтов во всем мире (СВО, столкновения на Ближнем Востоке и в Африке и др.) [2]. По данным экспертно-аналитического центра в 2023 г. доля внешних нарушителей выросла только в России [1]. В 2024 г. Российская Федерация заняла 2-е место по количеству утечек информации во всем мире, на первом месте оказались Соединенные штаты Америки.

Сегодня все чаще используются математические аппараты для описания различных процессов во всех областях деятельности человека [3]. Математические модели нашли широкое применение и при описании преступности [4]. С помощью математической модели удастся изучить динамику преступности, прогнозировать дальнейшее ее развитие, а главное выявлять влияющих на нее факторы [5]-[6].

Целью работы является подбор модели для описания количества скомпрометированных ПДн для ее прогноза на 2025–2026 гг.

Самой простой моделью для описания статистических данных является линейная модель, которой характерна следующая зависимость:

$$Y = a + b \cdot X = a + b \cdot (T - 2012),$$

где a , b – являются некоторыми константами, T – время (год). Константы подбирают с использованием метода наименьших квадратов. Коэффициент линейной корреляции при этом равен 0.775. На рисунке представлены результаты вычислений.

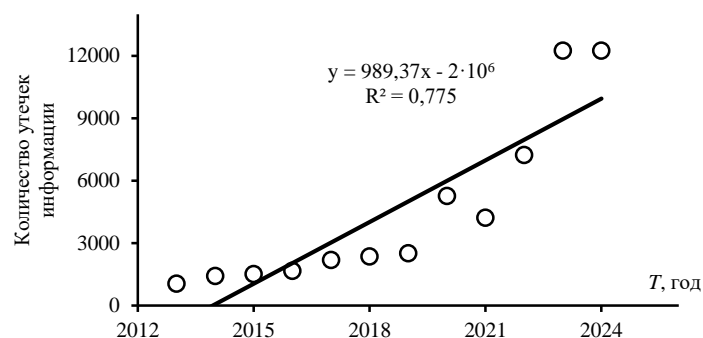


Рис. 1. Сравнение количества фактического числа утечек информации с линейной моделью (символ – статистические данные, линия – расчет)

В табл. 1 приведены результаты, полученные по различным классам математических функций, которые могут подойти к описанию имеющейся зависимости между Y (количества утечек) и T (год).

Таблица 1

Результаты проведенных вычислений

Модель	Вид уравнения	R^2
Линейная	$Y = a \cdot X + b$	0,775
Экспоненциальная	$Y = a \cdot \exp(b \cdot X)$	0,9407
Показательная	$Y = k \cdot a^{b \cdot X}$	0,9535
Степенная	$Y = a \cdot X^b$	0,8584
Гиперболическая	$Y = a + b/X$	–
Перла – Рида	$Y = k/(1 + a \cdot \exp(b \cdot X))$	–
Гомпертца	$Y = a \cdot \exp(b \cdot \exp(c \cdot X))$	0,0823
Гомпертца – Мейкхама	$Y = a \cdot \exp(b \cdot k^{c \cdot X} + d)$	0,0957

Описать статистические данные не удалось с помощью гиперболической моделью и зависимостью Перла-Рида в виду особенности функциональной зависимости. По результатам вычислений видно, что наибольший коэффициентом регрессии обладает показательная модель 0.9535, что свидетельствует о тесной связи между расчетными и статистическими данными. Показательная функция применяется для аналитического описания тенденции развития показателя с течением времени [7]:

$$Y = k \cdot a^{b \cdot X},$$

где X = время–2012; K , a , b – некоторые константы.

Расчет коэффициентов показал следующие данные:

$$Y = 635 \cdot 1.08^{3.25 \cdot T}.$$

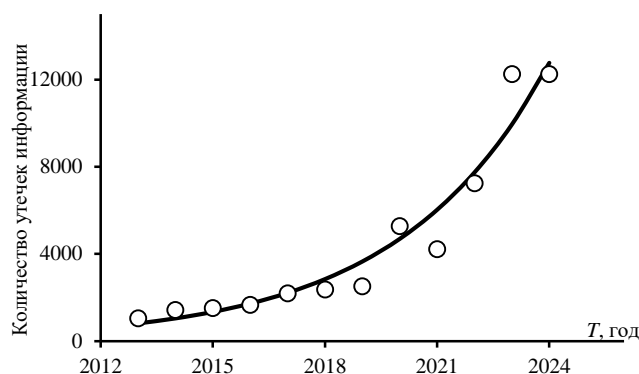


Рис. 2. Сравнение количества фактического числа утечек информации с показательной моделью (символ – статистические данные, линия – расчет)

При подстановке в полученное уравнение T , равное 2025–2027, получены следующие результаты:

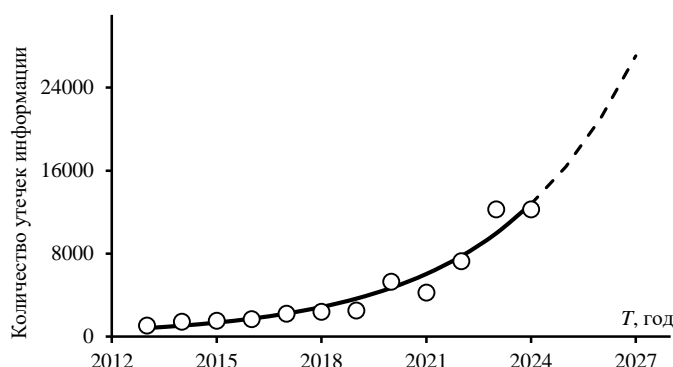


Рис. 3. Прогноз количества числа утечек информации (символ – статистические данные, линия – расчет)

Таким образом, несмотря на все объективные проблемы рынка информационной безопасности, многим крупным компаниям удастся планомерно выстраивать эффективные системы безопасности и поддерживать их соответствие новым угрозам. Несмотря на это в настоящий момент следует ожидать увеличение количества утечек информации в 2025 г. в 1,5 раза по сравнению с 2024 г. В марте Роскомнадзор опубликовал информацию об утечке 24 млн записей россиян с начала 2025 г. [8].

Список использованных источников:

1. Утечки информации в мире, 2022–2023 гг. // Экспертно-Аналитический центр InfoWatch. 2024 г. URL: <https://www.infowatch.ru/sites/default/files/analytics/files/issledovaniye-utechek-informatsii-v-mire-za-2022-2023-gody.pdf?ysclid=m9kwhev8ln209074874> (дата обращения: 17.04.2025).
2. Количество слитых персональных данных в 2024 г. выросло на треть // Платформа для бизнес-планирования. 05.03.2025 г. URL:

https://safe.cnews.ru/news/line/2025-03-05_kolichestvo_slityh_personalnyh
(дата обращения: 17.04.2025).

3. Галимнуров А.А. Применение математического моделирования для прогнозирования спреда корпоративных облигаций / А.А. Галимнуров, А.С. Исмагилова // Инженерный вестник Дона. 2025. № 2(122). С. 721–737.

4. Постовалов О.В. Математическое моделирование на основе теории причинности в приложении к криминалистической характеристике и профилактике преступлений // Правовое государство: теория и практика. 2020. С. 152–160.

5. Судакова Т.М. Математическое моделирование в разработке и совершенствовании методологических основ криминологии // Уголовно-правовые науки. 2023. Т. 24. № 3. С. 388–395.

6. Дианов Д.В. Прикладные статистические исследования экономической безопасности. М.: Русайнс. 2022. 194 с.

7. Сизова Т.М. Статистика для бакалавров. Часть II. СПб: Университет ИТМО, 2016. 70 с.

8. Роскомнадзор: в сеть утекло 24 млн записей россиян с начала 2025 г. Iaassaaspaas. Обзор ИТ решений для бизнеса и государства. 13.03.2025 г. URL: <https://iaassaaspaas.ru/news/okolo-milliarda-strok-personalnyh-dannyh-uteklo-v-set-za-polgoda-nikakie-shtrafy-ne-pomogayut> (дата обращения: 17.04.2025).

Kornilova A.A., Gizzatova E.R.

Ufa University of Science and Technology, Ufa

MATHEMATICAL MODELING OF THE TIME SERIES OF THE QUANTITY OF COMPROMISED PERSONAL DATA IN THE RUSSIAN FEDERATION

Abstract. The paper describes statistical data on the amount of compromised personal data in the Russian Federation in the period from 2013 to 2024. It is shown that the data are described by the exponential function with the highest correlation coefficient. An attempt to predict the number of information leaks for 2025 is given.

Keywords: mathematical modeling, forecast, personal data, leak.