Министерство науки и высшего образования Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Институт кибербезопасности и защиты информации

Работа допущена к защите
Директор Института
кибербезопасности и защиты
информации, д.т.н., проф.
_____ Д.П. Зегжда
«___»_____2021 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

РАБОТА БАКАЛАВРА

ОБУЧАЮЩИЙ КОМПЛЕКС ПО ДИФФЕРЕНЦИАЛЬНОЙ КОНФИДЕНЦИАЛЬНОСТИ ДЛЯ СПЕЦИАЛИСТОВ ПО ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

по направлению подготовки (специальности)
10.03.01 Информационная безопасность

Направленность (профиль)
10.03.01_03 Безопасность компьютерных систем

Выполнил
студент гр. 4831001/70301 В.С. Липатникова

Руководитель
доцент ИКиЗИ,
к.т.н., доцент
М.А. Полтавцева

Санкт-Петербург
2021

РЕФЕРАТ

На 63 с., 24 рисунка, 5 таблиц.

КЛЮЧЕВЫЕ СЛОВА: ДИФФЕРЕНЦИАЛЬНАЯ КОНФИДЕНЦИАЛЬНОСТЬ, МЕТОД ОБЕСПЕЧЕНИЯ ПРИВАТНОСТИ ДАННЫХ, МЕХАНИЗМ ЛАПЛАСА, РАНДОМИЗИРОВАННЫЙ АЛГОРИТМ, МЕХАНИЗМ ГАУССА

Выпускная квалификационная работа на тему «Обучающий комплекс по дифференциальной конфиденциальности для специалистов по информационной безопасности» направленна на теоретическое и практическое понимание дифференциальной конфиденциальности.

Целью работы является повышение качества подготовки специалистов по информационной безопасности путем разработки обучающего комплекса по обеспечению дифференциальной конфиденциальности. Задачами, которые решались в процессе работы, являются:

– проведение обзора существующих методов обеспечения конфиденциальности данных;

– анализ механизмов обеспечения дифференциальной конфиденциальности;

– проектирование обучающего комплекса по дифференциальной конфиденциальности для студентов по информационной безопасности.

В качестве объекта исследования выступают основные свойства и базовые механизмы обеспечения дифференциальной конфиденциальности. Предметом исследования являются механизмы Лапласа, Гаусса, а также рандомизированный и экспоненциальный алгоритмы.

Составленные в результате лабораторные работы содержат задания по программной реализации описанных в теоретической части алгоритмов, вычисления ошибок при применении этих алгоритмов с выявлением причин их возникновения, упражнения для лучшего понимания математического аппарата и основ дифференциальной конфиденциальности, а также контрольные вопросы для проверки знаний, полученных в ходе выполнения лабораторных работ.

Результаты выпускной работы могут быть полезны для студентов, обучающихся информационной безопасности, а также для людей, заинтересованных в теме обеспечения конфиденциальности данных.

THE ABSTRACT

63 pages, 24 figures, 5 tables.

KEY WORDS: DIFFERENTIAL PRIVACY, DATA PRIVACY METHOD, LAPLACE MECHANISM, RANDOMIZED ALGORITHM, GAUSSIAN MECHANISM

The final qualification work on the topic "Training complex on differential privacy for information security specialists" is aimed at theoretical and practical understanding of differential privacy.

The aim of the work is to improve the quality of training of information security specialists by developing a training complex for ensuring differential confidentiality. The tasks that were solved in the course of work are:

- review of existing data privacy practices;
- analysis of differential privacy mechanisms;
- designing a training complex on differential privacy for students on information security.

The object of the study is the main properties and basic mechanisms for ensuring differential confidentiality. The subject of the study is the Laplace and Gaussian mechanisms, as well as randomized and exponential algorithms.

The resulting laboratory work contains tasks for the software implementation of the algorithms described in the theoretical part, calculations of errors in the application of these algorithms with the identification of the causes of their occurrence, exercises for a better understanding of the mathematical apparatus and the basics of differential confidentiality, as well as control questions to test the knowledge gained during the laboratory work.

The results of the final work can be useful for students studying information security, as well as for people interested in the topic of data privacy.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В современности любая информация, которая становится известна о человеке, неважно был ли это сбор статистики, анкетирование населения, выборы или просто выполнение запросов под своей учетной записью, для аналитиков может стать кладом, который может раскрыть конфиденциальную информацию человека. С одной стороны подобный сбор информации о людях полезен с точки зрения улучшения качества жизни, однако, с другой стороны, эти данные могут содержать частную информацию, которой пользователь не хочет делиться с общественностью. Так, например, ученые компании Microsoft продемонстрировали, что, анализируя большие выборки запросов поисковых систем, в некоторых случаях они могут идентифицировать интернет-пользователей, страдающих раком поджелудочной железы, даже до того, как им будет поставлен диагноз этого заболевания [1].

Следовательно, с развитием технологий, которые обеспечивают все более детальный и эффективный сбор информации, возрастает потребность в надежном и математически строгом определении конфиденциальности, а также в обширном вычислительном классе алгоритмов, удовлетворяющих этому требованию.

Для определения режима конфиденциальности общедоступной информации в России в 2006 году был утвержден федеральный закон №152 «О персональных данных», где в статье 3 «Основные понятия, используемые в настоящем Федеральном законе», пункт 8 появляется понятие «обезличивание персональных данных» [2].

Приказ Роскомнадзора №996 Российской Федерации [3] устанавливает методологии обезличивания персональных данных, где описываются такие методы как: метод введения идентификаторов, метод изменения состава или семантики, метод декомпозиции, метод перемешивания. Однако последние уже не являются столь актуальными. В связи с этим многие компании обращаются

к иным, более новым методам, среди которых можно выделить дифференциальную конфиденциальность.

Актуальностью исследования является распространения понимания дифференциальной конфиденциальности среди специалистов по информационной безопасности с целью дальнейшего развития и применения этого метода обеспечения приватности данных, при которой сохраняется баланс между сохранением полезной нагрузки информации и ее конфиденциальности.

В качестве *объекта исследования* данной работы являются фундаментальные свойства и основные механизмы обеспечения дифференциальной приватности, а *предметом исследования* являются экспоненциальный механизм и механизмы Лапласа и Гаусса, которые обеспечивают приватность набора данных.

Целью этой выпускной квалификационной работы является повышение качества подготовки специалистов информационной безопасности путем разработки обучающего комплекса по обеспечению дифференциальной конфиденциальности.

Для достижения вышеописанной цели были поставлены следующие задачи:

– провести обзор существующих методов, подходов и алгоритмов обеспечения конфиденциальности данных, а также извлечь их преимущества и недостатки;

– проанализировать механизмы обеспечения дифференциальной конфиденциальности;

– спроектировать обучающий комплекс по дифференциальной конфиденциальности для студентов информационной безопасности.

В качестве *теоретической основы* использовались исследования Дворк С. и Рота А., а также Фатхима Ш., Дефонтейна Д. и др. Практическая часть была выполнена на основании документации по языку программирования Python, блокнота Jupyter Notebook, и СУБД Oracle.

Информационной базой исследования послужили знания и данные, полученные в результате прохождения курсов «Системы управления базами данных», «Теория вероятностей и математическая статистика», а также зарубежные статьи [18-22] и публикации [13-16,23] на тему дифференциальной конфиденциальности.

Научная новизна и *практическая значимость* настоящей работы заключается в анализе методов анонимизации данных, выделении базовых подходов и реализации нового обучающего комплекса для специалистов информационной безопасности по обеспечению дифференциальной конфиденциальности.

1. АНАЛИЗ МЕТОДОВ ОБЕСПЕЧЕНИЯ АНОНИМНОСТИ ПЕРСОНАЛЬНЫХ ДАННЫХ

В последние десятилетия сбор информации различными компаниями и организациями стал обычным явлением. Полученные данные связаны с определенной потребностью и относятся к какой-либо конкретной области. Как пример, правительственные организации могут собирать информацию в области медицины или экономики для оценки качества жизни населения. Большинство собранных данных при этом персонализированы, а значит, содержат частную информацию.

Для обеспечения конфиденциальности было предложено множество различных методов, подходов и алгоритмов сохранения конфиденциальности, среди которых можно выделить:

- анонимизация;
- метод добавления шума на основе возмущений;
- метод добавления мультипликативного шума на основе возмущений;
- метод K-анонимизации;
- методы на основе криптографии;
- дифференциальные механизмы конфиденциальности.

1.1 Анонимизация

Данные не могут быть полностью анонимными и оставаться полезными. Чем большей полнотой и структурированностью обладают данные, тем они интереснее и полезнее. Для решения данной проблемы ввели такие понятия, как «анонимность» и «удаление идентификационной информации лиц», которые обеспечивают безвозвратную потерю связности между данными и конкретным субъектом. Однако, если у злоумышленника есть вспомогательная информация из других источников, то анонимизации оказывается недостаточно, так как путем сопоставления этих данных с теми, что были анонимизированы, есть возможность идентификации пользователя.

Это доказывают исследования Латаня Суини [4, 5], в которых было показано, что даже при удалении всех явных идентификаторов отдельные лица все равно могут быть повторно идентифицированы посредством связей с другим общедоступным набором данных с помощью комбинации почтового индекса, даты рождения и пола.

Методы подобные анонимизации также не обладают композициональностью. Это значит, что при многократном обращении к данным с целью их анализа может появиться проблема потери конфиденциальности и приватности.

1.2 Метод добавления шума на основе возмущений

Метод добавления мультипликативного шума на основе возмущений, обеспечивает конфиденциальность и целостность информации. Данный метод достаточно эффективен с математической точки зрения [6-8]. Преимуществом такого подхода является возможность обеспечения конфиденциальности данных во время их сбора. Это обосновывается тем, что дальнейший процесс анализа данных никак не зависит от количества добавляемых возмущений. Но в то же самое время данный плюс обращается в минус, так как снижается общая защищенность информации при искажении. Еще одним недостатком этого метода является то, что последний подвержен атаке ввода-вывода, когда нарушителю известно какое-то количество информации и измененная шумом версия этих же данных. Используя реверс-инжиниринг, атакующих может узнать природу преобразований, которые произошли с данными.

1.3 Добавления мультипликативного шума на основе возмущений

Различные методы добавления шума [9, 10], являются с одной стороны эффективными, так как добавление возмущений увеличивает защищенность данных, но с другой стороны из-за применения шума можно потерять информацию вовсе. Из-за этого возникает необходимость поиска компромисса между попытками защиты информации и сохранения ее полезности.

1.4 Метод k -анонимизации

Метод k-анонимизации [4], является самой популярной моделью конфиденциальности и считается многообещающим, так как использует методы подавления и обобщения информации, которые позволяют добиться отсутствия конфиденциальной информации.

Весь набор данных разбивается на несколько групп – классов эквивалентности, в которых каждая запись имеет одинаковые значения атрибутов как минимум с k-1 другими записями. Однако при этом появляется проблема выбора данных для обезличивания, так как, при обезличивании всех значащих полей может возникнуть снижение информативности, а при обезличивании только части – возникает ситуация переизбытка данных в полях. Также этот метод подвержен атакам, в которых при помощи ассоциаций между атрибутами злоумышленник может получить защищаемые дынные.

1.5 Методы на основе криптографии

Применяя криптографические протоколы, можно создать анонимные каналы сообщений. Ишай, Кушелевиц и Островский [11] показывают, что анонимные каналы могут повысить эффективность для некоторых вычислительных задач. Существенным недостатком методов, основанных на криптографии, является подверженность внутренним атакам на каналы. Хоть системы и в состоянии отследить злоумышленника, они не могут предотвратить его деятельность, направленную на нарушение конфиденциальности, во время общения участников канала. В результате чего многие схемы имеют низкую эффективность и высокую стоимость.

1.6 Дифференциальные механизмы конфиденциальности

Все вышеизложенные методы не обладают твердой теоретической основой для оценки уровня конфиденциальности набора данных. Дифференциальная конфиденциальность в отличие них предлагает строгое определение конфиденциальности и набор технологий, удовлетворяющий этому определению.

Особенностью этого метода является то, что он обеспечивает анонимность пользователей, позволяя при этом проводить значимый анализ набора данных. При попытке злоумышленника или аналитика обратиться к базам данных, которые отличаются только одной записью, вероятность изменения конечного результата не будет зависеть от наличия, отсутствия или модификации этой записи, то есть при использовании дифференциальной конфиденциальности невозможно различить базы данных на основе результатов запроса. Таким образом, метод защищает личную информацию человека независимо от того, какая другая информация может быть доступна потенциальному злоумышленнику.

Еще одним существенным преимуществом дифференциальной конфиденциальности над ранее рассмотренными методами является то, что она обладает устойчивостью к композициональным атакам. Это значит, что обеспечивается защита от проблем потери приватности, возникающих при многократном обращении к данным с целью их анализа.

Описываемый метод привлекателен для множества приложений, поскольку гарантирует «потребности» конфиденциальности, то есть при этом подходе исключаются любые потенциальные методы, которые могут понадобиться аналитику данных, чтобы отличить конкретного человека от других участников или связать конкретное поведение с ранее идентифицированным человеком. Следовательно, дифференциальная конфиденциальность следит за тем, чтобы риск конфиденциальности, связанный с наличием информации о пользователе в базе, существенно не увеличивался [12].

***

Таким образом, в данном разделе выпускной квалификационной работы был произведен анализ существующих методов обеспечения анонимности, в результате которого были рассмотрены уязвимости, связанные с нарушением конфиденциальности субъекта. Также были исследованы достоинства и

недостатки вышеизложенных методов. Результаты анализа представлены в таблице 1.

Таблица 1 – Сравнительная таблица существующих методов обеспечения анонимности

Метод	Оценочные параметры алгоритмов, сохраняющих конфиденциальность		Недостатки	Преимущества
	Уровень конфиденциальности	Качество данных после применения метода		
Метод добавления шума на основе возмущений	Энтропия	Первоначальные данные могут быть раскрыты с высокой точностью	Подвержен атаке на основе известного образца, что позволяет предсказывать значения данных	Позволяет избежать разглашения данных и нарушения их целостности.
Метод добавления мультипликативного шума на основе возмущений	Разница дисперсий между фактическим и возмущенным значением	Точные значения исходного массива данных найти невозможно, но приближенные исходные данные можно	Подвержен атаке ввода-вывода.	Конфиденциальность данных может быть обеспечена во время процесса их сбора.
Метод K-анонимизации	Зависит от степени случайности (как минимум, 1/K)	Первоначальные данные могут быть раскрыты.	Подвержен к атакам с однородностями, если нарушитель обладает первоначальной информацией о конфиденциальном множестве данных	С помощью обобщения и усечения детализация представления данных понижается. И в конечном множестве данных отсутствует конфиденциальная информация.

Таблица 1 – Сравнительная таблица существующих методов обеспечения анонимности (продолжение)

Метод	Оценочные параметры алгоритмов, сохраняющих конфиденциальность		Недостатки	Преимущества
	Уровень конфиденциальности	Качество данных после применения метода		
Методы на основе криптографии	Зависит от криптографических методов.	Высокий уровень конфиденциальности.	Плохая эффективность.	Хорошо могут обеспечить конфиденциальность данных.
Дифференциальные механизмы конфиденциальности	Зависит от формы шумов (может быть Гауссовский шум или шум Лапласа)	Высокий уровень конфиденциальности	Ответы на запросы ограничены, из-за чего работа с БД останавливается для предотвращения утечки информации.	Защита здесь гарантируется, даже когда атакующий направляет серию адаптивных запросов к БД, изменяя количество шума, добавляемого к каждому результату запроса.

2. ОБЕСПЕЧЕНИЕ АНОНИМНОСТИ ДАННЫХ ПУТЕМ ДИФФЕРЕНЦИАЛЬНОЙ КОНФИДЕНЦИАЛЬНОСТИ

Основная цель дифференциальной конфиденциальности – гарантировать, что различные виды статистического анализа не нарушат конфиденциальность данных. То есть необходимо убедиться в том, что в результате применения статистического анализа к набору данных, содержащему в себе какую-либо информацию о человеке, не будет поставлена под угрозу конфиденциальность любого конкретного лица, содержащегося в этом наборе данных.

Конфиденциальность сохраняется, если после анализа анализатор ничего не знает о людях в наборе данных. Еще говорят, что они остаются «незамеченными».

Определим понятие «конфиденциальности». В 1977 году было предложено следующее определение: «Всю информацию, которую можно узнать об участнике в статистической базы данных, можно узнать и без доступа к этой базе данных» [13]. Это говорит о том, что все данные, которые мы хотим получить о человеке, должны быть только публичными. Однако по этому определению, во-первых, совершенно нельзя будет выборочно использовать информацию из каких-либо «частных» баз данных или источников, а это говорит о том, что эти данные перестают нести в себе смысл их использования в целом. А, во-вторых, появляется необходимость следить за тем, являются данные публичными или нет. Все это привело к новому определению, которое сформулировала Синтия Дворк: дифференциальная конфиденциальность описывает обещание, которое дает владелец информации субъекту данных, которое звучит следующим образом: «Вы никак не пострадаете, если позволите использовать ваши данные в каком-либо исследовании или анализе, независимо от того, какие еще наборы данных или источники информации доступны» [14]. Последняя часть содержит одну из ключевых идей, лежащих в основе дифференцированной конфиденциальности и дающих преимущество: она направлена на защиту личной информации человека независимо от того, какая другая информация

может быть доступна потенциальному злоумышленнику (например, в других наборах данных).

Рассмотрим рисунок 1, где представлен некоторый процесс, на вход принимающий базу данных, а на выходе возвращающий некоторые значения.

Рисунок 1 – Демонстрация получения процессом данных

Дифференциальная приватность — это свойство, которое обеспечивает конфиденциальность процесса, путем внесения случайности в набор данных. Значит для того, чтобы сделать процесс дифференциально частным, необходимо применить рандомизированный алгоритм, который будет вносить в изначальные данные некую меру случайности или, как еще называют, шум.

Дифференциальная конфиденциальность достигается, когда злоумышленник не может отличить ответы, полученные рандомизированным алгоритмом на запрос от изначальной базы данных и базы данных без одного человека, как показано на рисунке 2.

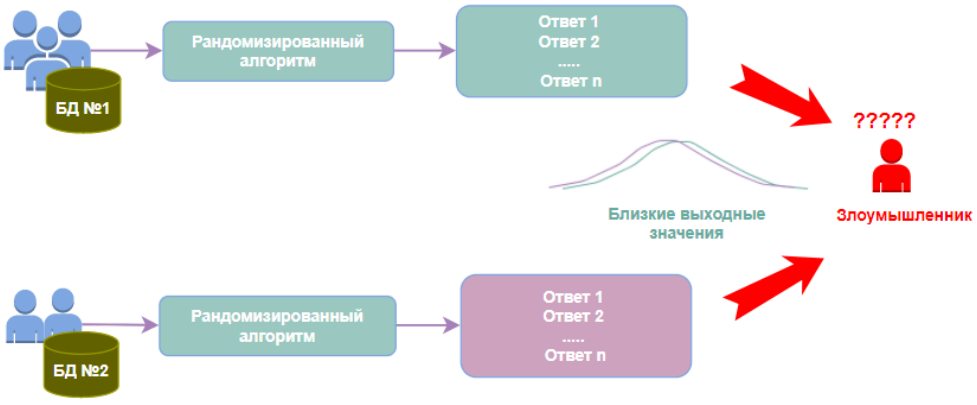

Рисунок 2 – Неразличимость баз данных злоумышленником

То есть изучаемый метод позволяет человеку находиться в базе данных и дает гарантию, что злоумышленник не сможет об этом узнать. Если распределения вероятностей будут близки друг к другу, то влияние одного человека на любой результат из возможных будет незначительным. Это

позволяет людям участвовать в опросах или исследованиях и предоставлять личную информацию, так как есть уверенность в том, что их данные буду защищены.

2.1 Типы анализа

Обработка данных подразумевает под собой различные виды анализа информации. Рассмотрим методы, для которых обеспечивается дифференциальная конфиденциальность.

1. Численные запросы – запросы, направленные на оценку количества отдельных записей в базе данных, которые удовлетворяют определенному свойству. Например, это может быть запрос числа людей в университете с голубыми глазами.
2. Гистограммы содержат информацию, классифицированную по непересекающимся категориям.
3. Кумулятивные функции распределения отражают вероятность того, что результат запроса будет меньше заданного значения или равен ему.
4. Линейная регрессия, которая показывает зависимость одной переменной от другой. Например, исследователь может попытаться понять, как здоровье человека зависит от его образования и дохода.
5. Кластеризация – метод анализа, включающий группировку точек данных в кластеры по общим признакам.
6. Классификация определяет категорию, к которой относятся данные.
7. Синтетические данные – это данные, полученные в результате ручного создания данных или путем применения какого-либо генератора [15].

2.2 Формальное определение

Введем формальное определение дифференциальной конфиденциальности, которая гарантирует, что рандомизированный алгоритм будет выдавать близкие выходные значения при схожих входных данных.

Пусть $\mathcal{E}$ – положительное действительное число, A – вероятностный алгоритм, который принимает на вход набор данных, а $P -$ вероятность.

Определение 1:

Базы данных будем называть *соседними*, если они отличаются только наличием или отсутствием одного человека.

Определение 2:

Рандомизированный алгоритм A обеспечивает (ε, δ) – *дифференциальную конфиденциальность,* если для соседних баз данных *D1* и *D2* и для всех $S \subseteq Range(A)$ выполняется следующее выражение:

$$P[A(D1) \in S] \leq exp^{\varepsilon} * P[A(D2) \in S] + \delta,$$

где ε называют *бюджетом приватности*, а параметр δ характеризует *вероятность случайной утечки информации*.

Если значение параметра $\mathcal{E}$ близко к 0, то экспоненциальная функция exp^{ε} близка к 1, а значит вероятности приблизительно одинаковы. Тогда достигается полная конфиденциальность данных, однако получается, что точность алгоритма низкая. Это вызвано тем, что к данным прибавляется большое количество шума. Следовательно, чем больше значение $\mathcal{E}$, тем сильнее вероятности отличаются друг от друга, то есть информация менее конфиденциальна.

Определение 3:

Если $\delta = 0$, то алгоритм A называют ε – *дифференциально конфиденциальным* и говорят, что для всех соседних баз данных *D1* и *D2* и всех выходных значений алгоритма A нельзя различить, какая база данных является истинной, на основе наблюдения за выходными данными.

Важно отметить, что дифференциальная конфиденциальность – это свойство не данных, а алгоритма A, который обеспечивает распределение вероятностей в диапазоне $Range(A)$.

2.3 Свойства дифференциальной конфиденциальности

Дифференциальная конфиденциальность обладает тремя основными свойствами:

1. Последовательной композицией.
2. Параллельной композицией.

3. Устойчивостью к постобработке.

Рассмотрим каждое более подробно.

2.3.1 Последовательная композиция

Последовательная композиция ограничивает конфиденциальность при публикации нескольких результатов дифференциально частных механизмов, полученных на одном и том же наборе данных. По теореме о последовательной композиции [16]: если $A_1(D)$ обладает ε_1 – дифференциальной конфиденциальностью, а $A_2(D)$ обладает ε_2 – дифференциальной конфиденциальностью, то механизм $A(D) = (A_1(D), A_2(D))$, который выводит оба результата, удовлетворяет $\varepsilon_1 + \varepsilon_2$ – дифференциальной конфиденциальности.

Пусть $A_1(D)$ обладает бюджетом конфиденциальности $\varepsilon_1 = 1$, а $A_2(D)$ имеет $\varepsilon_2 = 2$. Продемонстрируем приватность, используя механизм Лапласа.

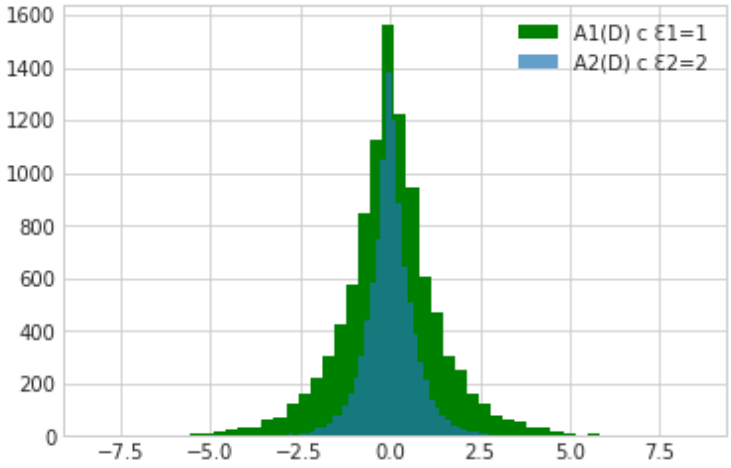

Рисунок 3 – Плотность распределений Лапласа для дифференциально частных алгоритмов $A_1(D)$ и $A_2(D)$

Как видно из рисунка 3 $A_2(D)$ обеспечивает большую вероятность получения результатов, близких к действительному значению и, следовательно, меньшую приватность.

Рассмотрим графики плотностей последовательной композиции $A_1(D)$ и $A_2(D)$.

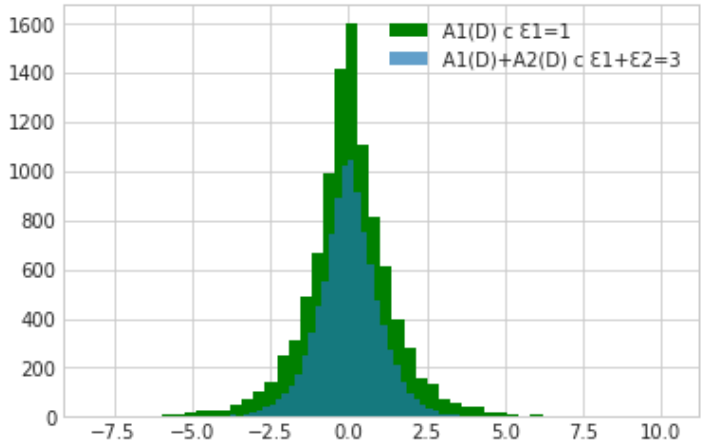

Рисунок 4 – Сравнение графиков плотности вероятности $A_1(D)$ и последовательной композиции $A_1(D)$ и $A_2(D)$

Рисунок 4 показывает, что последовательная композиция дает более точные выходные значения, что неудивительно, так как бюджет конфиденциальности $\varepsilon_1 + \varepsilon_2$ больше ε_1. Однако рисунок 5 изображает, что рассматриваемая композиция обеспечивает больший разброс величин. Это связано с тем, что конфиденциальность здесь описывается последовательным составом и, так как ε_1 в $\varepsilon_1 + \varepsilon_2$ меньше ε_2 в $A_2(D)$, то уровень приватности выше.

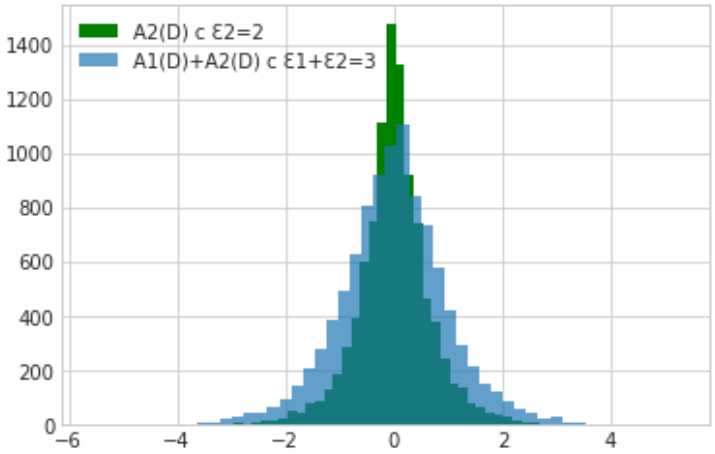

Рисунок 5 – Сравнение графиков плотности вероятности $A_2(D)$ и последовательной композиции $A_1(D)$ и $A_2(D)$

Таким образом, последовательная композиция позволяет контролировать уровень конфиденциальности, устанавливая границы бюджета приватности, что добавляет еще одно преимущество в копилку дифференциальной конфиденциальности.

2.3.2 Параллельная композиция

Параллельная композиция базируется на идее разделения набора данных D на непересекающиеся множества $D = d_1 \cup d_2 \cup \dots \cup d_k$ с дальнейшим применением механизма дифференциальной конфиденциальности $A(D)$ к каждому из этих множеств. После этого говорят, что результаты $A(d_1), \dots, A(d_k)$ удовлетворяют ε - дифференциальной конфиденциальности.

Важно сказать, что параллельная композиция обеспечивает лучшую приватность, так как рандомизированный механизм выполняется для каждого подмножества только один раз, что дает ε - дифференциальную конфиденциальность. Последовательная же композиция же складывается из бюджетов конфиденциальности каждого механизма $\varepsilon_1 + \dots + \varepsilon_k$, а это удовлетворяет $k * \varepsilon$ – дифференциальной приватности.

Примером применения параллельной композиции являются гистограммы, разбивающие набор данных на основе значения одного из атрибутов, как показано на рисунке 6.

2.3.3 Постобработка

Свойство постобработки говорит о том, что если некоторый рандомизированный алгоритм $A(D)$, применяемый к набору данных D, удовлетворяет ε – дифференциальной конфиденциальности, то для любой (детерминированной или рандомизированной) функции g выполняется $g(A(D))$, которое обеспечивает ε – дифференциальную приватность.

Иными словами, постобработка позволяет выполнять произвольные вычисления с данными, полученными в результате применения к ним дифференциально частного механизма, не опасаясь при этом за потерю конфиденциальности.

Рассматриваемое свойство применяют для повышения точности или уменьшения шума дифференциально частных алгоритмов, а также используется для гарантии ограничения ε – дифференциальной конфиденциальности при

наличии у злоумышленника или стороннего лица вспомогательной информации, которая может содержаться в функции g.

Курс	Истинное число студентов	Зашумленное число студентов
1	20	22.412667289183837
2	24	23.84715523342759
3	20	18.32513345143855
4	31	29.513973857037797
5	27	26.36052101224095
6	25	25.264189299295502

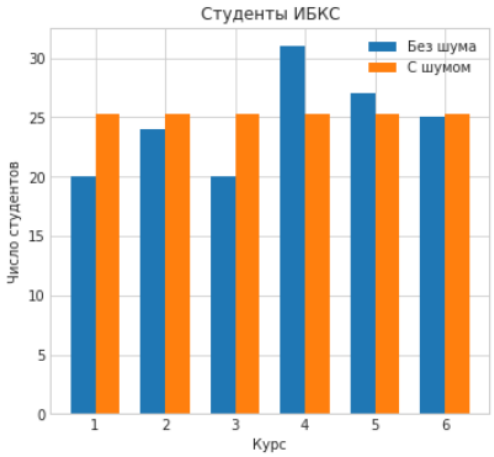

Рисунок 6 - Пример применения параллельной композиции

2.3.4 Расширенная композиция

В определении 2 дифференциальной конфиденциальности указывался параметр δ, обозначающий вероятность возникновения ошибки при работе рандомизированного алгоритма.

Для (ε, δ) – дифференциальной конфиденциальности существует дополнительное свойство, касающееся композиции – расширенная композиция. Оно основана на идее наличия последовательности механизмов $m_1, \dots, m_k$, где каждый m_i может быть получен из результатов всех предыдущих механизмов $m_1, \dots, m_{i-1}$. Основная теорема расширенной композиции гласит, что если механизм m_i является ε – дифференциально конфиденциальным, то общий бюджет конфиденциальности при $\delta \geq 0$ может быть получен как:

$$\varepsilon_{\text{общий}} = 2 * \varepsilon \sqrt{2 * k * \log\left(\frac{1}{\delta}\right)}$$

Композиция позволяет оценить верхние границы значения Ɛ, то есть сколько приватной информации можно потерять при нескольких запросах с разными значениями Ɛ.

2.4 Количественная оценка знаний злоумышленника

Предположим, что к базе данных *D* применен алгоритм *A,* который вносит случайность. Цель злоумышленника – выяснить, находится ли определенный человек в базе данных.

Рассмотрим худший случай, когда злоумышленнику известна вся база данных, однако он не знает в какой из двух баз данных *D1* или *D2* находится интересующий его человек. Атакующий может сделать предположение $P[D = D1]$, которое будет означать, что его цель находится в базе данных *D1* с некоторой вероятностью $P \in [0,1]$. Точно таким же образом может быть выдвинуто подозрение, что в *D2* цель злоумышленника отсутствует $P[D = D2] = 1 - P[D = D1]$.

Предположим, что после применения *A* к базе данных возвращается выходное значение *Out*, исходя из которого атакующий может сделать новое предположение, которое определяется как $P[D = D2 \mid A(D) = Out]$. Применяя теорему Байеса, получаем:

$$P[D = D2 \mid A(D) = Out] = \frac{P[D = D2\,] * P[\,A(D) = Out | D = D2\,]}{P[\,A(D) = Out]},$$

где $P[D = D2\,]$ - изначальное предположение злоумышленника, $P[\,A(D) = Out | D = D2\,]$– вероятность получения значения *Out* из базы данных *D2*, $P[\,A(D) = Out]$ – вероятность того, что в результате применения алгоритма *A* к *D2* выходным значением будет *Out.*

Если первоначальным предположением злоумышленника окажется, что нужные ему данные содержатся в *D1*, тогда аналогичным образом получится:

$$P[D = D1 \mid A(D) = Out] = \frac{P[D = D1\,] * P[\,A(D) = Out | D = D1\,]}{P[\,A(D) = Out]}$$

Так как значение $P[\,A(D) = Out]$ неизвестно, нужно произведи преобразование, чтобы эта величина сократилась, а для этого рассмотрим соотношение:

$$\frac{P[D = D2 \mid A(D) = Out]}{P[D = D1 \mid A(D) = Out]} = \frac{P[D = D2]}{P[D = D1]} * \frac{P[\,A(D2) = Out]}{P[\,A(D1) = Out]} \qquad (1)$$

Применяя к $P[\,A(D2) = Out]$ и $P[\,A(D1) = Out]$ определение дифференциальной конфиденциальности, получим:

$$\frac{P[\,A(D2) = Out]}{P[\,A(D1) = Out]} \leq \ exp^{\varepsilon}$$

Так как при перестановке баз данных они все равно будут отличаться одной записью, выражение выше можно модифицировать:

$$exp^{-\varepsilon} \leq \frac{P[\,A(D2) = Out]}{P[\,A(D1) = Out]} \leq \ exp^{\varepsilon} \qquad (2)$$

Из (1) и (2) выводим:

$$exp^{-\varepsilon} * \frac{P[D = D2]}{P[D = D1]} \leq \frac{P[D = D2 \mid A(D) = Out]}{P[D = D1 \mid A(D) = Out]} \leq \ exp^{\varepsilon} * \frac{P[D = D2]}{P[D = D1]}$$

Заменяя $P[D = D1]$ на $1 - P[D = D2]$, а также проделывая это для $P[D = D1 \mid A(D) = Out]$, получаем неравенство вида:

$$\frac{P[D = D2]}{exp^{\varepsilon} + (1 - exp^{\varepsilon}) * P[D = D2]} \leq P[D = D2 \mid A(D) = Out]$$

$$\leq \ exp^{\varepsilon} * \frac{P[D = D2]}{1 + (exp^{\varepsilon} - 1) * P[D = D2]}$$

$\mathcal{E}$ - дифференциальная конфиденциальность гарантирует, что для каждого применения алгоритма A результат приблизительно с равной вероятностью будет одновременно наблюдаться на одной из соседних баз данных [17]. Это значит, что статистические данные запросов и функций, выполняемых в базе, не должны сильно зависеть от конкретного человека.

Выведенное неравенство, на основе которого был построен график на рисунке 7, показывает количественную оценку этого явления.

Черная линия на графике означает, что атакующий практически не получает никакой информации, следовательно его предположения практически

не обновляются в силу недостатка данных. Цветные линии одинаковых цветов – это верхние и нижние границы знаний злоумышленника. Если Ɛ=0, тогда будет достигаться полная конфиденциальность людей, чья информация содержится в базе данных. При увеличении же значения знания злоумышленника о содержимом базы данных увеличивается, соответственно приватность будет уменьшаться и может снизиться до полного отсутствия.

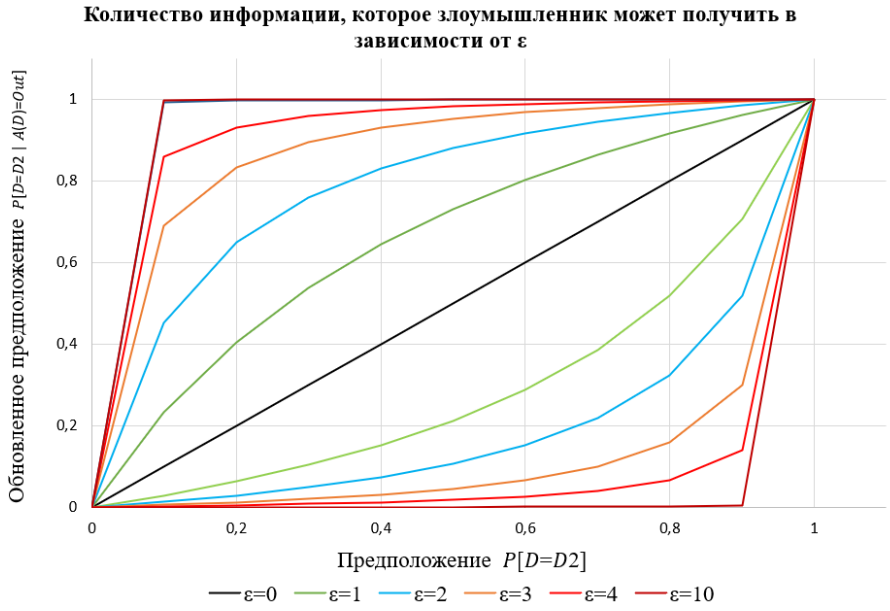

Рисунок 7 – График, отображающий количественную оценку $\mathcal{E}$ - дифференциальной конфиденциальности

2.5 Модели дифференциальной конфиденциальности

Как выяснили ранее, дифференциальная конфиденциальность добавляет шум к данным. В зависимости от того, где именно он добавляется выделяют два типа дифференциальной конфиденциальности – локальная и глобальная.

2.5.1 Глобальная модель

В глобальной модели, схема работы которой представлена на рисунке 8, каждый участник доверяет службе, которая собирает информацию, поэтому данные от пользователей отправляются без добавления шума. Рандомизированный алгоритм применяется один раз, когда служба публикует результаты или отвечает на запрос, полученный от третьих лиц.

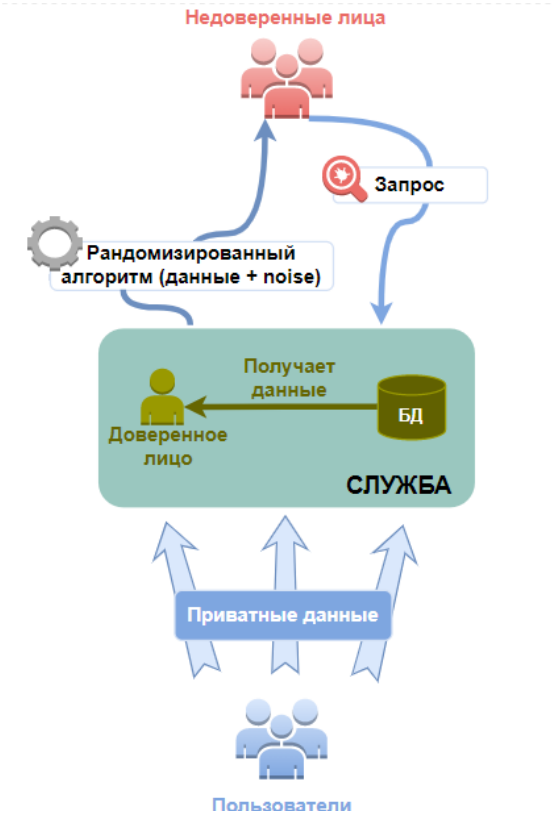

Рисунок 8 – Глобальная модель

Преимуществом модели является точность, так как в одном месте собираются данные о разных пользователях. В связи с этим каждый отдельный пользователь «теряется в толпе», что означает, что его частные данные извлечь становится сложнее, поэтому количество шума, которое необходимо добавлять мало. Однако можно выделить и недостатки, связанные с тем, что участники должны достаточно доверять службе, потому что последняя должна корректно добавлять шум и защищать конфиденциальность пользователя. Кроме того, в глобальной модели вся информация собирается в одном месте, что увеличивает риск отказа.

2.5.2 Локальная модель

В локальной модели, изображенной на рисунке 9, каждый пользователь перед предоставлением информации какой-либо службе применяет к своим данным рандомизированный алгоритм.

В отличие от предыдущего типа дифференциальной конфиденциальности, служба может сразу же публиковать полученную информацию без

дополнительных преобразований. То есть плюсом локальной модели является то, что служба не требует доверия, так как каждый участник сам защищает свои данные. Из минусов можно выделить то, что общий шум, наложенный на информацию, становится больше, а это значит, что точность уменьшается и для получения полезных результатов потребуется гораздо больше пользователей.

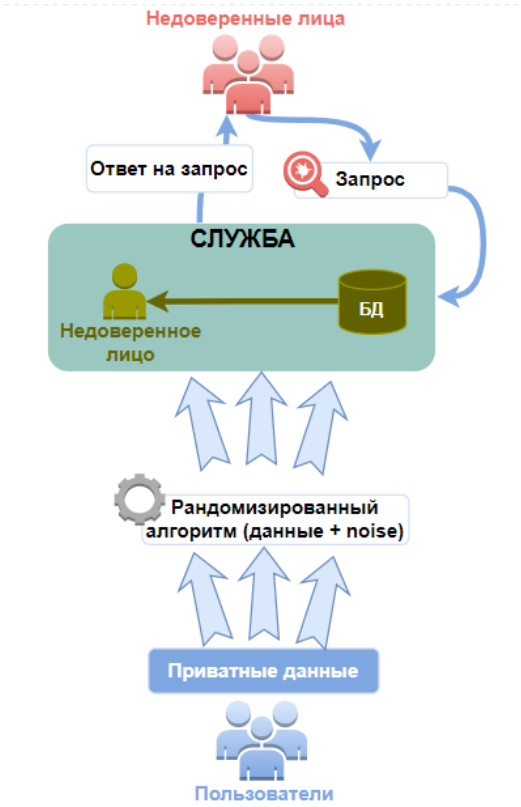

Рисунок 9 – Локальная модель

2.6 Чувствительность

Чувствительность запроса помогает понять, насколько данные человека влияют на расчеты и, следовательно какое количество шума необходимо добавлять. Высокая чувствительность Δf при фиксированном значении Ɛ служит предупреждением о том, что необходимо добавить больше шума, чтобы замаскировать данные [18]. Для дальнейшего рассмотрения чувствительностей введем определения норм *L1* и *L2*.

Определение 4:

Норма L1 вектора *V* длины *k* определяется как сумма элементов вектора:

$$\left|\left|V\right|\right|_1 = \sum_{i=1}^{k} |\mathrm{V_i}|$$

Определение 5:

Норма L2 вектора *V* длины *k* определяется как квадратный корень из суммы квадратов:

$$\left|\left|V\right|\right|_2 = \sqrt{\sum_{i=1}^{k} \mathrm{V_i^2}}$$

Существует глобальная чувствительность и локальная.

2.6.1 Глобальная чувствительность

Определение 6:

Глобальная чувствительность характеризуется максимальной разницей выходных данных, к которой может привести функция запроса f при внесении одного изменения в любой из всех возможных наборов данных, отличающихся не более чем одним элементом.

$$\Delta f_{GS} = \max_{D1,D2:d(D1,D2)=1} \left|\left|f(D1) - f(D2)\right|\right|_1,$$

где $\left|\left|.\right|\right|_1$ – расстояние L_1 – нормы между наборами данных, отличающихся не более, чем на один элемент [19].

Важно отметить, что глобальная чувствительность — это максимальная разница в выходных данных с учетом всех возможных наборов данных, и поэтому она зависит только от запроса, а не от набора данных.

Эта мера чувствительности называется «глобальной», потому что она не зависит от фактического запрашиваемого набора данных, она верна для любого выбора соседних *D1 и D2* [20].

2.6.2 Локальная чувствительность

Локальная чувствительность характеризуется максимальной разницей выходных данных, к которой может привести функция запроса f при внесении одного изменения в локальный набор данных *D1:*

$$\Delta f_{LS} = \max_{D2} \left|\left| f(D1) - f(D2) \right|\right|_1$$

Локальная чувствительность является функцией как запроса f, так и фактического набора данных $D1$. В отличие от случая глобальной чувствительности, нельзя говорить о локальной чувствительности функции, не учитывая также набор данных, в котором возникает эта локальная чувствительность, а значит, что известен и размер набора. Следовательно, есть возможность устанавливать конечные границы чувствительности некоторых функций.

Идея, лежащая в основе локальной чувствительности, заключается в том, чтобы зафиксировать один из двух наборов данных как фактический запрашиваемый набор данных и учесть всех его соседей. Глобальная чувствительность учитывает любые два соседних набора данных.

Таким образом, локальная чувствительность — это минимальная чувствительность, необходимая для того, чтобы запрос охватил один конкретный набор данных, а глобальная — это минимальная чувствительность, необходимая для того, чтобы запрос охватил все возможные наборы данных.

2.7 Механизмы добавления шума

Как было сказано ранее, дифференциальная конфиденциальность обеспечивает приватность данных путем внесения случайности или шума. Количество добавляемого шума зависит от четырех параметров:

- типа шума (Лапласа, Гаусса или экспоненциальный);
- чувствительности запроса или функции;
- желаемого Ɛ;
- желаемого δ.

Рассмотрим каждый из дифференциально частных механизмов добавления шума.

2.7.1 Механизм Лапласа

Механизм Лапласа используют для выполнения числовых запросов к базе данных. Эти запросы отображают базы данных на множество действительных чисел:

$$f: D \to R \text{ или } f: D \to R^k,$$

где R^k – вектор действительных чисел длины k.

Для определения механизма Лапласа требуется ввести следующее определение.

Определение 7:

Распределением Лапласа (с центром в точке 0) называется непрерывное распределение случайной величины с функцией плотности вероятности [21]:

$$Lap(f(D)|b) = \frac{1}{2 * b} * \exp\left(-\frac{|f(D)|}{b}\right),$$

где b – параметр масштаба, который определяет статистическую дисперсию распределения вероятностей или, иначе говоря, величина показывает, на сколько распределение «плоское».

Определение 8:

Механизм Лапласа заключается в добавлении шума – значений распределения Лапласа к истинному ответу на запрос $f(D)$:

$$A(D) = f(D) + Lap(b),$$

причем параметр $b = \frac{\Delta f}{\varepsilon}$, где Δf – чувствительность запроса, которая показывает сколько нужно добавлять шума к данным, чтобы они были конфиденциальные. Чем больше чувствительность, тем больше шума необходимо добавлять, чтобы обеспечить конфиденциальность информации.

Механизм Лапласа обеспечивает ε-дифференциальную конфиденциальность, так как из определений дифференциальной конфиденциальности и механизма Лапласа следует, что:

$$\frac{P[\,A(D) = Out]}{P[\,A(D`) = Out]} = \frac{\frac{1}{2*b} * exp\left(-\frac{|f(D)-Out|}{b}\right)}{\frac{1}{2*b} * exp\left(-\frac{|f(D`)-Out|}{b}\right)} = exp\left(\frac{|f(D)-Out|}{b}+\frac{|f(D`)-Out|}{b}\right)$$

$$= exp\left(\frac{\varepsilon*(|f(D`)-Out|-|f(D)-Out|)}{\Delta f}\right)$$

$$\leq exp\left(\frac{\varepsilon*|f(D`)-f(D)|}{\Delta f}\right) = exp\left(\frac{\varepsilon*||f(D`)-f(D)||_1}{\Delta f}\right) \leq exp^{\varepsilon}$$

Относительная ошибка вычисляется как разница между истинным результатом запроса и тем, что получился в результате зашумления:

$$error = value_{real} - value_{differential\ privacy}$$

Эта величина зависит от значения параметра ε в механизме Лапласа и отражает точность результата. Для $\varepsilon = 0{,}1$ относительная ошибка примерно в 10 раз больше, чем для $\varepsilon = 1{,}1$, что говорит о том, что ошибка обратно пропорциональна ε. Чем выше бюджет конфиденциальности, тем меньше ошибка, и наоборот. Это отражают гистограммы на рисунках 10 и 11, которые были построены по запросу, который вычисляет число людей в наборе данных, кто никогда не был в браке, зашумленному механизмом Лапласа для $\varepsilon = 0{,}1$ и $\varepsilon = 1{,}1$ соответственно.

При малых значениях ε чувствительности, равной 1, параметр $b = \frac{\Delta f}{\varepsilon} = \frac{1}{\varepsilon}$ будет большим, а шум, который будет наложен на результат запроса – нулевым. График плотности вероятности механизма Лапласа для запроса, ответ которого равен 500, изображенный на рисунке 12, показывает, что, чем меньше значение ε, тем более конфиденциальны данные. Красная гистограмма отображает значения ответа при $\varepsilon = 10$, серая – при $\varepsilon = 1$, зеленая – при $\varepsilon = 0{,}3$.

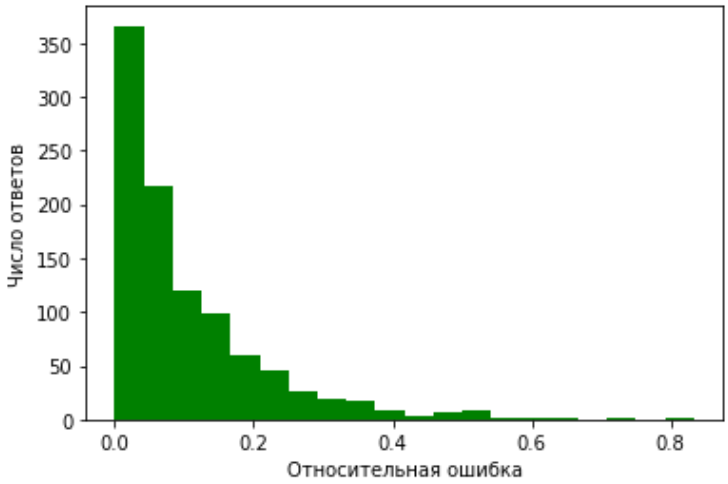

Рисунок 10– Гистограмма относительной ошибки для $\varepsilon = 0{,}1$

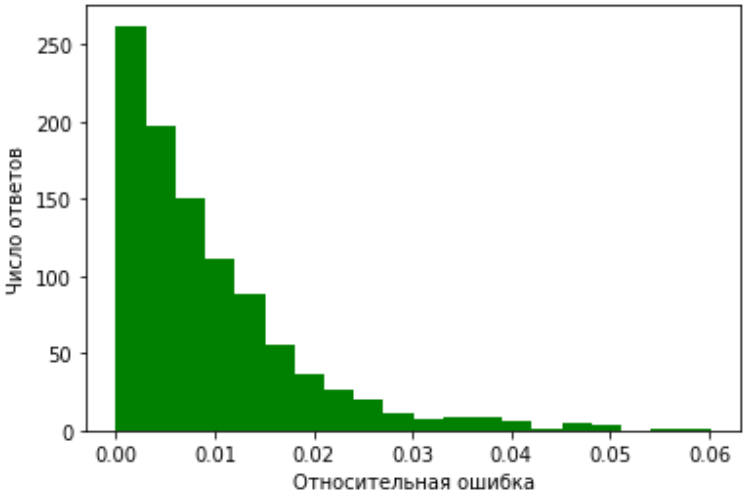

Рисунок 11 – Гистограмма относительной ошибки для $\varepsilon = 1{,}1$

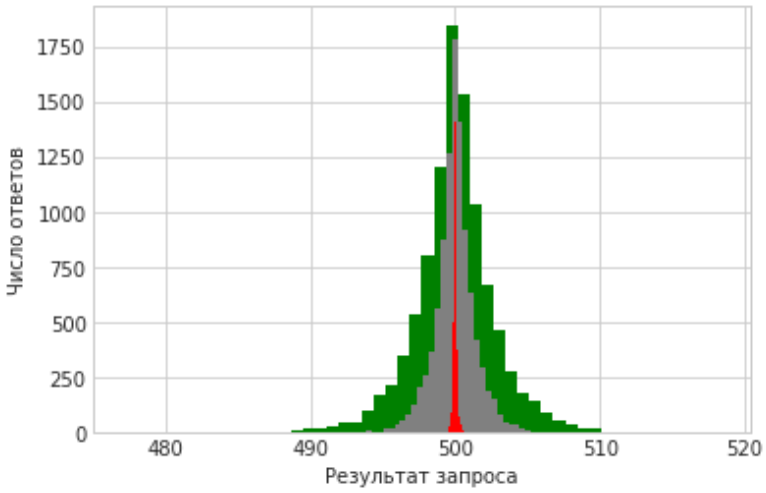

Рисунок 12 – Плотность вероятности для механизма Лапласа

2.7.2 Механизм Гаусса

В определении 2 был указан параметр δ, который обозначает вероятность нарушения корректной работы дифференциально частного механизма. То есть с вероятностью δ дифференциальная конфиденциальность будет нарушена, и с вероятностью $1 - \delta$ последняя будет установлена, а именно будет соблюдаться:

$$\frac{P[\,A(D2) = Out]}{P[\,A(D1) = Out]} \leq exp^{\varepsilon}$$

Механизмом, в котором наблюдается (ε, δ) – дифференциальная конфиденциальность, является механизм Гаусса.

Определение 8:

Механизм Гаусса заключается в добавлении шума – значений нормального распределения к истинному ответу на запрос $f(D)$ [22]:

$$A(D) = f(D) + \mathcal{N}(\sigma^2), \text{где } \sigma^2 = \frac{2 * \Delta f^2 * \log\frac{1.25}{\delta}}{\varepsilon^2}$$

Механизм схож с механизмом Лапласа, однако по сравнению с последним, при использовании механизма Гаусса вероятность получения результата, который не является истинным, выше, так как распределение более пологое, что показано на рисунке 13. Это говорит о том, что вероятность получения точных данных при использовании механизма Гаусса меньше.

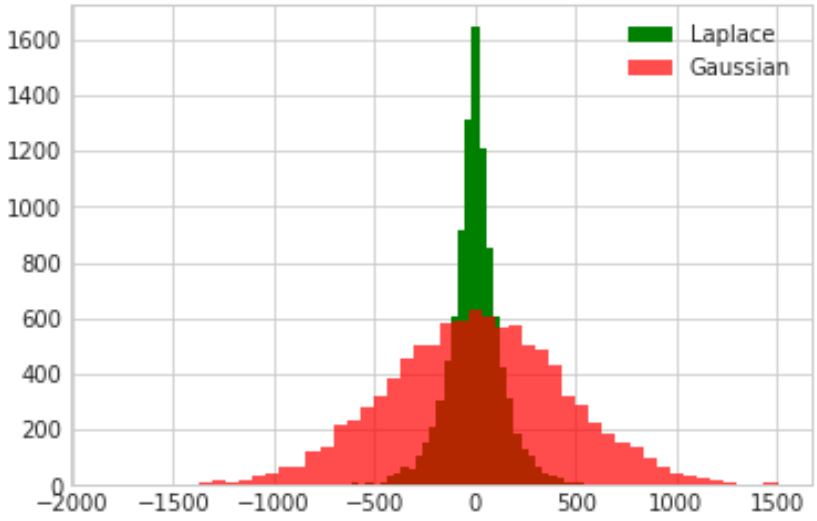

Рисунок 13 – Сравнение функций плотности вероятности механизмов Гаусса и Лапласа для $\varepsilon = 0.1, \delta = 10^{-5}$

Несмотря на такие недостатки как меньшая точность и вероятность δ нарушения дифференциальной конфиденциальности, механизм Гаусса по сравнению с механизмом Лапласа позволяет добавлять к данным гораздо меньше шума, так как может использовать чувствительность с нормой L2, которая меньше чувствительности с нормой L1 [1]. Предположим, что есть набор данных, в котором содержится информация о пациентах и специализация врача, которого данный пациент посещал. Каждая запись соответствует одному человеку. К набору данных D выполняется запрос $f(D)$, подсчитывающий число пациентов по каждой из специальностей врачей. Наглядно это можно изобразить на рисунке 14. Пусть число специалистов будет равняться 20. Важно сказать, что пациент может обратиться к нескольким врачам, однако если он посетит «терапевта» 5 раз, то количество пациентов «терапевта» увеличится на 1. Чтобы обеспечить дифференциальную приватность информации необходимо добавить шум в соответствии с чувствительностью.

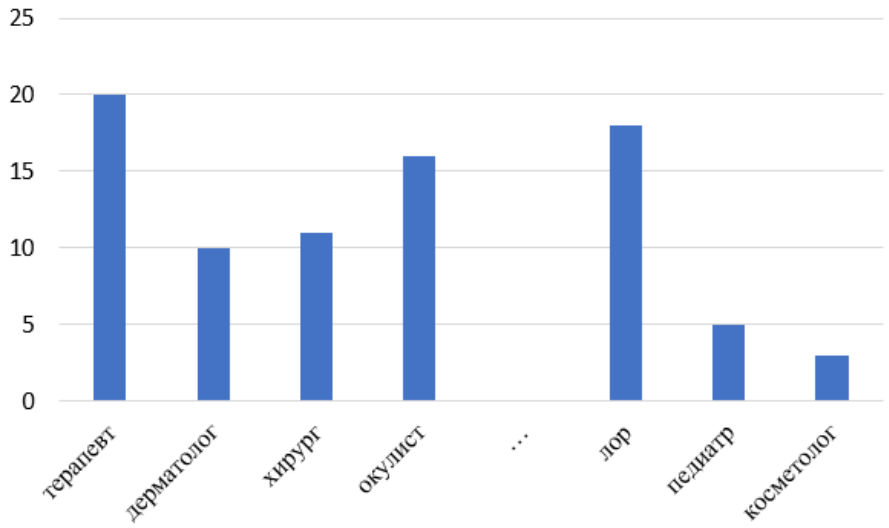

Рисунок 14 – Наглядное изображения выполнения запросов по каждому врачу

Чувствительность измеряет, насколько ее вывод может измениться при добавлении одной записи в базу данных. Как было сказано ранее, пациент изменяет статистику лишь на 1, значит по определению чувствительности из раздела 2.5 данной выпускной квалификационной работы, для механизма Лапласа получаем следующее значение чувствительности:

$$\Delta_1 f = \max_{D1,D2:d(D1,D2)=1} \sum_{i=1}^{20} |f(D1) - f(D2)| = 20$$

Для механизма Гаусса это значение будет равняться:

$$\Delta_2 f = \max_{D1,D2:d(D1,D2)=1} \sqrt{\sum_{i=1}^{20} |f(D1) - f(D2)|^2} = \sqrt{\sum_{i=1}^{20} 1^2} = \sqrt{20} \approx 4.47$$

Полученные результаты говорят о том, что в соответствии с меньшим значением чувствительности, механизм Гаусса добавляет к данным меньше шума, что хорошо, так как нам нужно обеспечить конфиденциальность информации, при этом сохраняя как можно большую точность.

Механизм Гаусса отлично подходит для обеспечения дифференциальной конфиденциальности при выполнении запросов к набору данных, где один человек оказывает влияние на множество величин. В противном случае, когда пользователь меняет только одно значение, лучше использовать метод Лапласа.

2.7.3 Экспоненциальный механизм

Экспоненциальный механизм используется для обеспечения дифференциальной конфиденциальности, когда выходные данные не являются числовыми. Механизм был разработан для обстоятельств, в которых было желательно выбрать наилучший ответ [23].

Пусть D, как и раньше, обозначает входной набор данных, а $r \in R$ обозначает один из потенциальных ответов на запрос, учитывая функцию оценки $u: D \times R \to \mathbb{R}$, случайный алгоритм A удовлетворяет ε-дифференциальной конфиденциальности, если выбирает ответ на основе вероятности:

$$A(D,u) = \left\{ r \middle| P[r \in R] = exp^{\frac{\varepsilon * u(D,r)}{2\Delta u}} \right\},$$

где Δu означает чувствительность функции оценки и определяется как:

$$\Delta u = \max_{r \in \mathbb{R}} \max_{D1,D2:||D1-D2||_1 \leq 1} |u(D1,r) - u(D2,r)|$$

Цель механизма состоит в том, чтобы случайным образом сопоставить набор данных D с некоторым выходом в диапазоне R, в результате чего любой

паре (d, r) присваивается оценка [24]. Максимальное значение этой оценки и будет результатом опроса. К примеру, некая компания хочет установить цену на выпускаемое программное обеспечение. Чтобы сделать это, они проводят опрос потенциальных покупателей, спрашивая цену, которую те готовы заплатить. При этом людям обещают, что названные каждым из них, не будут разглашены, опубликуется лишь дифференциально конфиденциальное значение оптимальной стоимости. Однако, если руководствоваться механизмами Лапласа или Гаусса, к цене необходимо будет добавить шум для обеспечения дифференциальной приватности, что изменит значение цены и может оказаться таковым, что люди не захотят столько платить, а это приведет к нулевым продажам. Экспоненциальный механизм решает эту проблему. Учитывая параметр конфиденциальности Ɛ, набор результатов R и функцию оценки u, которая сопоставляет пары - (база данных, потенциальный результат) с оценкой, механизм выбирает один элемент из R на основе распределения вероятностей, указанного выше. Вышеизложенный пример представлен на рисунке 15.

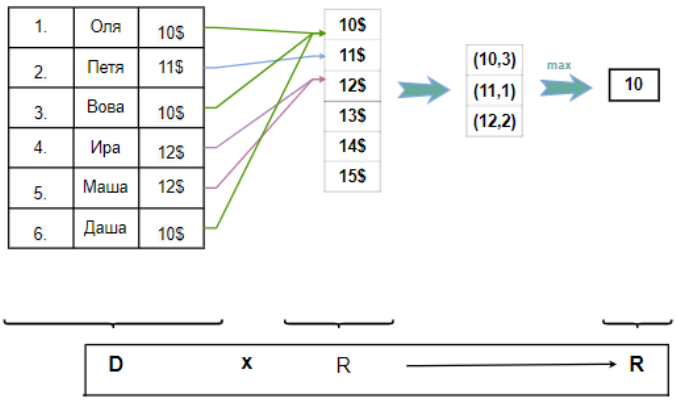

Рисунок 15 – Принцип работы экспоненциального механизма

При этом экспоненциальный механизм обеспечивает Ɛ - дифференциальную конфиденциальность:

$$\frac{P[\,A(D1,u)=r]}{P[\,A(D2,u)=r]}=\frac{\left(\frac{exp^{\frac{\varepsilon*u(D1,r)}{2\Delta u}}}{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D1,r`)}{2\Delta u}}}\right)}{\left(\frac{exp^{\frac{\varepsilon*u(D2,r)}{2\Delta u}}}{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D2,r`)}{2\Delta u}}}\right)}$$

$$=\left(\frac{exp^{\frac{\varepsilon*u(D1,r)}{2\Delta u}}}{exp^{\frac{\varepsilon*u(D2,r)}{2\Delta u}}}\right)*\left(\frac{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D2,r`)}{2\Delta u}}}{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D1,r`)}{2\Delta u}}}\right)$$

$$=exp^{\frac{\varepsilon*(u(D1,r)-u(D2,r))}{2\Delta u}}*\left(\frac{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D2,r`)}{2\Delta u}}}{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D1,r`)}{2\Delta u}}}\right)$$

$$\leq exp^{\frac{\varepsilon}{2}}*exp^{\frac{\varepsilon}{2}}*\left(\frac{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D1,r`)}{2\Delta u}}}{\sum_{r`\in R} exp^{\frac{\varepsilon*u(D1,r`)}{2\Delta u}}}\right)=exp^{\varepsilon}$$

***

Таким образом, в этой части работы были даны основные определения, играющие важную роль для обеспечения дифференциальной приватности алгоритмов. Также было рассмотрено формальное определение понятия дифференциальная конфиденциальность, его основные свойства, механизмы и особенности их применения. Кроме прочего был произведен анализ подсчета бюджета конфиденциальности ε, который позволяет произвести количественную оценку знаний злоумышленника.

3. РАЗРАБОТКА ЛАБОРАТОРНЫХ РАБОТ ДЛЯ СПЕЦИАЛИСТОВ ПО ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В этой части будет описана практическая часть выпускной квалификационной работы, которая посвящена созданию лабораторных работ по обеспечению дифференциальной конфиденциальности для студентов, обучающихся на направлении информационной безопасности.

3.1 Разработка структуры и тем лабораторных работ

В качестве тем для лабораторных работ были выбраны основные механизмы обеспечения дифференциальной конфиденциальности, которые в настоящее время используются известными компаниями для сохранения приватности информации пользователей. Среди таких механизмов можно выделить:

– механизм рандомизированного ответа, который применяется технологией RAPPOR компании Google для сбора статистических данных о населении, путем исследования запросов пользователей в поисковой строке [25];

– механизм Лапласа, используемый в системе PrivTree от Microsoft для маскировки местоположения людей в базах данных геолокации [26];

– механизм Гаусса, зачастую применяемый в сферах здравоохранения для конфиденциальности данных пациентов [27];

– экспоненциальный механизм, из которого вышеупомянутые методы могут быть получены [21].

Изучаемые алгоритмы обеспечения дифференциальной конфиденциальности позволят студентам освоить актуальную на сегодняшний день базу, которую в дальнейшем можно будет расширять, углубляясь в тему обеспечения приватности данных.

Структура каждой лабораторной работы представляет собой 2 части - теоретическую и практическую. Теоретическая часть составляет приблизительно 40% и включает в себя контрольные вопросы, ответы на которые можно дать, основываясь на содержащемся в лабораторной работе теоретическом материале,

а также упражнения, решение которых требует от студента не только понимание предоставленной информации, но и более глубокую оценку механизмов и свойств дифференциальной конфиденциальности. Упражнения включают в себя задачи на доказательство приватности, анализ и интерпретацию полученных результатов, а также в некоторых работах учащийся может представить себя в роли специалиста по дифференциальной приватности для обеспечения конфиденциальности набора данных и оценки степени его защищенности.

Практическая часть составляет 60% лабораторной работы. Она представляет собой реализацию механизмов и свойств дифференциальной приватности для выданного набора данных. От студента требуется также произвести анализ разработанных методов, а также продумать и реализовать способы их улучшения. Помимо прочего в этой части необходимо строить графики, гистограммы и сравнительные таблицы, основываясь на результатах, полученных после реализации запросов к набору данных, защищенному с помощью механизмов дифференциальной конфиденциальности.

Вышеописанная структура позволяет познакомить специалиста по информационной безопасности не только с основами рассматриваемого в выпускной работе метода обеспечения анонимности данных, но и закрепить полученные знания на практике посредствам проведения различного анализа и решения некоторого множества задач.

3.2 Инструменты реализации

Лабораторные работы, реализованные в рамках данной выпускной квалификационной работы, разработаны в интерактивном веб-приложении Jupyter Notebook [28], которое обеспечивает удобство и наглядность разработки. Его преимущества перед составлением лабораторных работ лишь в Microsoft Word отражены в таблице 2.

Использование скриптов также не является удобным, так как они не наглядны и могут лишь запутать студента.

Таблица 2 – Преимущества составления лабораторных работ с поддержкой Jupyter Notebook помимо Microsoft Word

Критерии сравнения	Jupyter Notebook	Microsoft Word
Содержимое предоставляемых шаблонов кода	Содержат меньше опечаток, а также намного легче в запуске кода.	Возможны опечатки в примерах, плюс ко всему для запуска кода необходимо его скопировать и, возможно, изменить внешний вид для корректности отступов и прочего.
Наглядность программной составляющей	Подсвечивание частей кода различными цветами обеспечивает более наглядную картину для понимания логики программы.	Весь код одного цвета, что делает его осмысление более сложным.
Проверочные тесты	Меньшая вероятность опечаток, уже имеет корректный синтаксис.	Могут также содержать опечатки или иметь некорректный синтаксис, что запутает учащегося.
Отправная точка	Студентам будет сразу понятно, где и что делать, без траты дополнительного времени на поиск среды реализации.	Среда выполнения кода не так очевидна, поэтому отнимет у учащегося время на ее поиски.
Входные наборы данных	Веб-приложение реализует удобную обработку наборов данных и имеет множество документации на эту тему.	
Составление графиков	Обеспечивает удобство в создании графиков, гистограмм, таблиц, схем.	

Таким образом, Jupyter Notebook позволяет совместить все необходимые составляющие – код, текст, изображения и графики в одном месте, благодаря чему обеспечить комфортную среду выполнения лабораторных работ.

Языком программирования, используемым для выполнения заданий, является Python. Этот выбор объясняется тем, что последний позволяет

подключать огромное количество библиотек, содержит большое число документации, а также является довольно простым и понятным языком.

3.3 Состав лабораторных работ

Каждая лабораторная работа посвящена отдельной теме, касающейся дифференциальной конфиденциальности, и содержит 7 вариантов для выполнения.

Состав лабораторных работ, который предоставляется студентам, выглядит следующим образом:

1. Блокнот Jupyter Notebook, который содержит:

– название лабораторной работы;

– цель лабораторной работы;

– теоретические сведения, которые содержат информацию, необходимую для выполнения практических задач;

– задания на работу по вариантам, где описывается последовательность задач, а также макет кода, графики, проверочные тесты по вариантам и пояснения, нужные для облегчения и большего понимания требований от студента;

– шаблоны, содержащие комментарий «#ВАШ КОД», для выполнения студентом программной части задания.

2. Документ формата .docx с дублированием названия, цели, теоретической информации и заданий, однако, помимо этого, включающий в себя:

– упражнения для более глубокого понимания материала. Они включают в себя задачи, касающиеся математического аппарата дифференциальной конфиденциальности; вопросы, поясняющие полученные в ходе работы результаты; головоломки на проверку студента в роли аналитика по дифференциальной конфиденциальности;

– контрольные вопросы, касающиеся базовых основ, ответы на которые можно найти в теоретической части;

– содержание отчета с требованиями для учащихся.

3. Раздаточный материал формата .csv с указанием вариантов, который содержит наборы данных, анализируемые в ходе выполнения лабораторных работ.

Для преподавателя дополнительно предоставляются:

1. Блокнот Jupyter Notebook, который содержит требуемые реализации задач.

2. Документ формата .docx, в состав которого плюсом прилагаются решения упражнений и ответы на контрольные вопросы.

3.3.1 Работа 1 – Механизм рандомизированного ответа

Целью работы является изучение механизма рандомизированного ответа и анализ результатов в зависимости от набора данных.

Рандомизированный ответ — это механизм локальной дифференциальной конфиденциальности, который изначально предназначался для улучшения предвзятости ответов на «тонкие» вопросы. Например, рассмотрим ситуацию, когда необходимо узнать зарплату сотрудников некоторой компании. Если напрямую спросить людей, то велика вероятность, что большинство не захотят сообщать запрошенную информацию, так как она относится к частной жизни. Выход из данной ситуации заключается в том, что каждый сотрудник прибавляет к своей заработной плате S случайное число R, находящееся в промежутке от -10000 до 10000, и сообщает в ответ на запрос лишь полученную в результате сумму $Summ = S + R$. Так как $Summ$ не выдает истинную зарплату человека, то можно сказать, что персональные данные были защищены. То есть для того, чтобы обеспечить конфиденциальность данных, полученных от некоторых лиц, дифференциальная конфиденциальность добавляет к исходной информации статистический шум, как показано на рисунке 16.

Однако, в таком случае, необходимо найти баланс между сохранением конфиденциальности данных и информативностью, так как после внесения случайности снизилась информативность данных. Для этого используется закон

простых чисел, благодаря которому по мере роста размера выборки ее среднее значение приближается к среднему, то есть шум устраняется и полученное среднее значение становится близко к истинному среднему числу. И чем больше выборка, тем больше будет приближение.

Рисунок 16 – Механизм работы дифференциальной конфиденциальности

Идея алгоритма состоит в том, что человек отвечает на вопрос в соответствии с результатом подбрасывания двух монет, как представлено на рисунке 17. Это вносит некую неопределенность, которая и обеспечивает конфиденциальность информации.

Алгоритм позволяет ввести степень правдоподобного отрицания, так как аналитик данных уже не может быть уверен в истинности ответа определенного человека. Однако, в то же время, достаточно информативен, чтобы позволить сделать какие-то выводы.

Рисунок 17 – Алгоритм рандомизированного ответа

В соответствии с номером варианта учащемуся предоставляется набор данных, как представлено в таблице 3, который содержит список людей с их личной информацией (идентификатор человека, имя, дата рождения, возраст, образование, профессия, семейное положение, пол, страна проживания, зарплата и др.). Пример набора данных представлен на рисунке 18.

Таблица 3 – Наборы данных в соответствии с вариантом студента

Вариант	Имя набора данных	Размер набора данных (КБ)
1	data_for_lab1_var1.csv data_for_lab1_var1_100.csv data_for_lab1_var1_1000.csv data_for_lab1_var1_10000.csv	4808 15 150 1492
2	data_for_lab1_var2.csv data_for_lab1_var2_100.csv data_for_lab1_var2_1000.csv data_for_lab1_var2_10000.csv	4813 15 150 1492
3	data_for_lab1_var3.csv data_for_lab1_var3_100.csv data_for_lab1_var3_1000.csv data_for_lab1_var3_10000.csv	4823 15 150 1492
4	data_for_lab1_var4.csv data_for_lab1_var4_100.csv data_for_lab1_var4_1000.csv data_for_lab1_var4_10000.csv	4834 15 150 1492
5	data_for_lab1_var5.csv data_for_lab1_var5_100.csv data_for_lab1_var5_1000.csv data_for_lab1_var5_10000.csv	4838 15 150 1492
6	data_for_lab1_var6.csv data_for_lab1_var6_100.csv data_for_lab1_var6_1000.csv data_for_lab1_var6_10000.csv	4853 15 150 1492
7	data_for_lab1_var7.csv data_for_lab1_var7_100.csv data_for_lab1_var7_1000.csv data_for_lab1_var7_10000.csv	4856 15 150 1492

Основными задачами лабораторной работы являются:

1. Осуществление запроса к набору данных data_for_lab1_varX.csv без использования рандомизированного алгоритма, значение которого будет считаться истинным ответом.

2. Применение алгоритма рандомизированного ответа к предоставленному набору данных.

3. Выполнение запроса к набору данных с использования рандомизированного алгоритма, значение которого будет считаться дифференциально частным ответом.

4. Вычисление относительной ошибки при применении механизма рандомизированного ответа.

5. Сравнение результатов ошибок при использовании предоставляемых наборов данных различных размеров – data_lab1_varX_100.csv, data_lab1_varX_1000.csv, data_lab1_varX_10000.csv.

6. Выполнение упражнений:

1) докажите, что механизм рандомизированного ответа является дифференциально частным;

2) механизм рандомизированного ответа является локальной моделью дифференциальной конфиденциальности. Объясните, в каких случаях следует использовать локальную модель, а в каких глобальную.

	Name	DOB	SSN	Zip	Age	Workclass	fnlwgt	Education	Education-Num	Marital Status	Occupation	Relationship	Race	Sex	Capital Gain
0	Dinnie Finding	3/26/1995	859-26-1365	47690	45	State-gov	252208	HS-grad	9	Separated	Adm-clerical	Own-child	White	Female	0
1	Enid Couttes	11/12/1990	720-61-8855	65093	41	NaN	202822	HS-grad	9	Separated	NaN	Not-in-family	Black	Female	0
2	Torrin Godon	11/17/2002	174-05-8234	48013	72	NaN	129912	HS-grad	9	Married-civ-spouse	NaN	Husband	White	Male	0
3	Ogdon Waren	6/6/1970	264-33-3644	80926	45	Local-gov	119199	Assoc-acdm	12	Divorced	Prof-specialty	Unmarried	White	Female	0
4	Jennine O'Hegertie	7/28/1953	544-48-2737	21613	31	Private	199655	Masters	14	Divorced	Other-service	Not-in-family	Other	Female	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
32535	Templeton Darwent	6/8/1955	167-13-0691	85917	22	Private	325033	12th	8	Never-married	Protective-serv	Own-child	Black	Male	0
32536	Andrew Hobbing	8/8/1990	157-53-5697	97686	34	Private	160216	Bachelors	13	Never-married	Exec-managerial	Not-in-family	White	Female	0
32537	Meredeth Pickavance	8/11/1962	531-67-5318	57639	30	Private	345898	HS-grad	9	Never-married	Craft-repair	Not-in-family	Black	Male	0
32538	Brande Tschierse	11/18/1982	218-94-3412	91234	38	Private	139180	Bachelors	13	Divorced	Prof-specialty	Unmarried	Black	Female	15020
32539	Lazar Dalgliesh	10/6/1976	235-28-4118	98532	71	NaN	287372	Doctorate	16	Married-civ-spouse	NaN	Husband	White	Male	0

Рисунок 18 – Набор данных для варианта 1

При выполнении работы результаты относительных ошибок для различных вариантов будут выглядеть приблизительно так, как представлено в таблице 4. Эти значения показывают, что механизм рандомизированного ответа хорошо выполняет свою задачу для большой базы данных, так как в этом случае по закону больших чисел выходные значения принимают приблизительно истинное значение, что обеспечивает приватность, и в то же самое время гарантирует точность результата. В случае маленького содержания строк в наборе, ответы на запросы перестают быть точными, из-за чего информативность содержимого теряется.

Таблица 4 – Приблизительные выходные значения при реализации лабораторной работы 1

Вариант	Относительная ошибка (в %) в наборах данных, содержащих:			
	32219 строк	100 строк	1000 строк	10000 строк
1	1,888	99,665	98,447	65,997
2	5,242	99,880	98,609	65,873
3	4,849	99,666	97,435	69,276
4	0,479	99,773	98,241	66,418
5	0,399	99,987	98,029	70,439
6	1,369	100	98,139	70,395
7	3,662	99,668	98,461	71,921

3.3.2 Работа 2 – Механизм Лапласа

В этой работе производится анализ исходного набора данных, конфиденциальность которого обеспечивается с помощью применения механизма Лапласа.

В отличие от механизма рандомизированного ответа механизм Лапласа позволяет оценивать количество генерируемого шума для обеспечения дифференциальной конфиденциальности. Также среди плюсов этого алгоритма можно отметить то, что он является более точным.

В качестве исходных данных используется информация из Национального отделения неотложной помощи, содержащая сведения об употреблении пациентами зловредных препаратов [29]. На рисунке 19 представлены первоначальные материалы.

Далее с помощью функции, представленной ниже производится генерация датасета, изображенный на рисунке 20, с которым проводится дальнейшая работа. Он содержит информацию о 100 людях и препаратах, которые они принимали.

```
        """Вход: data - фрейм данных, содержащий препарат и вероятность того, что пациент его использовал. N - количество людей, которых нужно сгенерировать. r - случайное значение.
        Выход: DATA - фрейм данных, содержащий N строк, где каждая строка соответствует человеку, 1 в столбце означает, что человек использовал лекарство."""
    def DatasetModeling(data, N, r = None):
        df = {}
        for index_num, row in data.iterrows():
            name = row['Substance']
            probability = row['Probability']
            df_row = bernoulli.rvs(probability, size=N, random_state = np.random.RandomState(seed=r))
            f[name] = df_row
            DATA = pd.DataFrame(df)
        return DATA
```

	Substance	Probability
0	Alcohol	0.530527
1	Cocaine	0.325744
2	Heroin	0.109529
3	Marijuana	0.218641
4	Stimulants	0.101317
5	Amphetamines	0.027766
6	Methamphetamine	0.077830
7	MDMA	0.007762
8	LSD	0.001503
9	PCP	0.027706
10	Antidepressants	0.094340
11	Antipsychotics	0.052264
12	Miscellaneous_hallucinogens	0.001879
13	Inhalants	0.009668
14	lithium	0.007860
15	Opiates	0.174578
16	Opiates_unspecified	0.032600
17	Narcotic_analgesics	0.147420
18	Buprenorphine	0.001014
19	Codeine	0.009444
20	Fentanyl	0.007097
21	Hydrocodone	0.053049
22	Methadone	0.029202
23	Morphine	0.011054
24	Oxycodone	0.051556
25	Ibuprofen	0.027810
26	Muscle_relaxants	0.032265

Рисунок 19 – Содержимое набора данных для выполнения лабораторной работы 2

	Alcohol	Cocaine	Heroin	Marijuana	Stimulants	Amphetamines	Methamphetamine	MDMA	LSD	PCP	...
0	1	0	0	0	0	0	0	0	0	0	...
1	1	0	0	0	0	0	0	0	0	0	...
2	1	0	0	0	0	0	0	0	0	0	...
3	0	1	0	1	0	0	0	0	0	0	...
4	1	0	0	0	0	0	0	0	0	0	...

Рисунок 20 – Сгенерированный набор данных

В лабораторной работе необходимо проанализировать среднее значение $\theta = E[X]$ по полученному материалу дифференцированно частным образом.

Основываясь на том, как были сгенерированы данные, истинное среднее значение распределения, из которого были взяты выборки, — это исходные вероятности из набора данных Национального отделения неотложной помощи, которое вычисляется как:

```
TRUE_MEAN = data['Probability'].to_numpy()
```

Затем требуется реализовать и сравнить результаты трех алгоритмов.

Алгоритм 1 – Не конфиденциальный

Алгоритм не является дифференциально конфиденциальным. Оценивать среднее значение необходимо путем определения среднего значения выборок D_i:

$$\bar{\theta} = A(D) = f(D) = \frac{1}{N}\sum_{i=1}^{N} D_i$$

Алгоритм 2 – Механизм Лапласа

Механизм Лапласа, который добавляет шум ε_e к истинному ответу на запрос $f(D)$ как:

$$\bar{\theta} = A(D) = f(D) + \varepsilon_e = \left(\frac{1}{N}\sum_{i=1}^{N} D_i\right) + \varepsilon_e$$

$\varepsilon_e \in R^d$ представляет собой независимые координаты, каждая из которых распределена в соответствии с параметром масштаба $\frac{\Delta f}{e}$. Чувствительность Δf определяется как:

$$\Delta f = \max_{D_1,D_2:d(D_1,D_2)=1} \left|\left| f(D_1) - f(D_2) \right|\right|_1 = \max_{D_1,D_2:d(D_1,D_2)=1} \left|\left| \frac{1}{N}\sum_{i=1}^{N} D_{1i} - \frac{1}{N}\sum_{i=1}^{N} D_{2i} \right|\right|_1$$

$$= \frac{d}{N}$$

где $\left|\left|.\right|\right|_1$ – расстояние L_1 – нормы.

Алгоритм 3 – Локально дифференциально конфиденциальный механизм Лапласа

В отличие от алгоритма 2 в данном алгоритме шум ε_e добавляется не к агрегированному значению $f(D)$, а к каждому D_i по отдельности:

$$\bar{\theta} = A(D) = f(D + \varepsilon_e) = \frac{1}{N}\sum_{i=1}^{N}(D_i + \varepsilon_e^i)$$

Здесь $\varepsilon_e \in R^{d\times N}$ делает каждую строку в данных дифференциально конфиденциальной еще до того, как будет получен ответ на запрос $f(D)$. Для каждой строки D_i, $\varepsilon_e^i \in R^d$ имеет независимые координаты, каждая из которых распределена в соответствии с распределением Лапласа с параметром масштаба $\frac{\Delta D_i}{e}$. Чувствительность ΔD_i изменения одной строки D_i базы данных определяется как:

$$\Delta D_i = \max_{D_1 \neq D_2} \left|\left| D_1 - D_2 \right|\right|_1 = d$$

где $\left|\left|.\right|\right|_1$ – расстояние L_1 – нормы.

Сравнение алгоритмов происходит путем оценки относительной ошибки каждого из них. Столбцы Alg1_error, Alg2_error, Alg3_error на рисунке 21 отражают разницу между значениями, полученными при выполнении алгоритмов, и настоящей величиной.

Среди трех алгоритмов только алгоритмы 2 и 3 обеспечивают дифференциальную конфиденциальность.

Наихудшая оценка наблюдается у алгоритма 3, так как ошибка в значениях больше двух предыдущих алгоритмов, что говорит о меньшей точности. Это связано с тем, что шум накладывается на каждую строку. Количество шума в

алгоритме 2 меньше за счет того, что значения «теряются в толпе», благодаря чему количество накладываемого шума уменьшается, и полезность данных остается на высоком уровне.

Алгоритм 3 может быть использован в случаях, когда человек не доверяет какой-либо службе, ответственной за обеспечение конфиденциальности, и накладывает на свои данные шум самостоятельно. Причинами недоверия могут быть как сомнения в качестве добавляемого шума или в человеке, отвечающем за наложение шума, так и опасение за то, что служба может быть взломана или атакована злоумышленником, из-за чего данные, собираемые последней, будут узнаны третьими лицами.

Out[314]:

	Substance	Probability	Algorithm1	Alg1_error	Algorithm2	Alg2_error	Algorithm3	Alg3_error
0	Alcohol	0.530527	0.527567	-0.002960	0.526615	-0.003912	0.563574	0.033047
1	Cocaine	0.325744	0.328867	0.003123	0.328980	0.003236	0.388357	0.062613
2	Heroin	0.109529	0.109433	-0.000095	0.107706	-0.001823	-0.866334	-0.975862
3	Marijuana	0.218641	0.220433	0.001793	0.216840	-0.001801	0.131663	-0.086977
4	Stimulants	0.101317	0.100267	-0.001051	0.101803	0.000485	0.198052	0.096735
5	Amphetamines	0.027766	0.027333	-0.000433	0.028008	0.000242	-0.269192	-0.296958
6	Methamphetamine	0.077830	0.077867	0.000037	0.078446	0.000616	-0.282818	-0.360648
7	MDMA	0.007762	0.007567	-0.000196	0.007862	0.000099	-0.046022	-0.053785
8	LSD	0.001503	0.001300	-0.000203	-0.003289	-0.004791	-0.279306	-0.280809
9	PCP	0.027706	0.027233	-0.000472	0.026631	-0.001075	0.558815	0.531109
10	Antidepressants	0.094340	0.093300	-0.001040	0.097296	0.002955	0.620563	0.526222
11	Antipsychotics	0.052264	0.052433	0.000170	0.048618	-0.003645	0.420631	0.368368
12	Miscellaneous_hallucinogens	0.001879	0.001667	-0.000213	0.004011	0.002132	-0.265626	-0.267505
13	Inhalants	0.009668	0.009333	-0.000335	0.011862	0.002194	-0.322983	-0.332651
14	lithium	0.007860	0.007700	-0.000160	0.003598	-0.004262	-0.312335	-0.320195
15	Opiates	0.174578	0.175867	0.001288	0.176521	0.001943	0.035884	-0.138695
16	Opiates_unspecified	0.032600	0.032133	-0.000466	0.032330	-0.000270	-0.349142	-0.381742
17	Narcotic_analgesics	0.147420	0.148033	0.000613	0.148424	0.001003	-0.666163	-0.813583
18	Buprenorphine	0.001014	0.000667	-0.000347	0.000606	-0.000408	-0.285246	-0.286260
19	Codeine	0.009444	0.009133	-0.000310	0.008109	-0.001335	-0.665541	-0.674984
20	Fentanyl	0.007097	0.006933	-0.000163	0.006000	-0.001097	0.871207	0.864110
21	Hydrocodone	0.053049	0.053233	0.000184	0.053470	0.000420	-0.362075	-0.415124
22	Methadone	0.029202	0.028600	-0.000602	0.028180	-0.001022	-0.315096	-0.344299
23	Morphine	0.011054	0.010500	-0.000554	0.012050	0.000996	0.144929	0.133875
24	Oxycodone	0.051556	0.051633	0.000077	0.051313	-0.000243	0.587558	0.536002
25	Ibuprofen	0.027810	0.027367	-0.000443	0.025125	-0.002685	-0.653888	-0.681697
26	Muscle_relaxants	0.032265	0.031933	-0.000332	0.029777	-0.002488	-0.129760	-0.162025

Рисунок 21 – Пример результатов выполненной лабораторной работы 2

3.3.3 Работа 3 – Механизм Гаусса и свойства дифференциальной конфиденциальности

Основная цель – изучение метода Гаусса и анализ основных свойств Ɛ и (Ɛ, δ) – дифференциальной конфиденциальности.

В данной лабораторной работе сначала требуется посчитать общее значение последовательной и расширенной композиции и сравнить полученные результаты. График, построенный в ходе выполнения этих задач, будет принимать вид, представленный на рисунке 22.

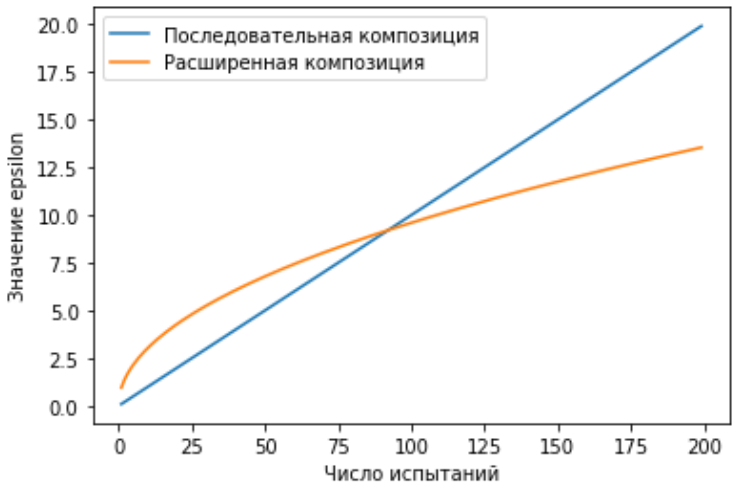

Рисунок 22 – Сравнение последовательной и расширенной композиций

Из вышеприведенного графика следует вывод о том, что при большом количестве испытаний (запросов) расширенная композиция обеспечивает более высокий уровень конфиденциальности, однако в случае малого числа запросов (в данном случае примерное менее 95) стоит отдать предпочтение последовательной композиции.

Оба графика выдают верные значения. Однако их результаты разные. Это происходит в силу того, что теорема о композиции задает верхнюю границу потери приватности, а не четкий уровень, т. е. на самом деле, бюджет конфиденциальности может быть и ниже этого максимума, но никак не выше.

Теоремы о композиции используются в случаях множественных запросов. Для обеспечения конфиденциальности необходимо наложить шум на результаты запросов. Поэтому следующим шагом лабораторной работы требуется реализовать векторный механизм Гаусса на основе примера реализации векторного механизма Лапласа, код которого выглядит следующим образом:

```
#Векторный механизм Лапласа
def LaplaceMechVec(vector, sensitivity, epsilon):
```

```
    return [vec + np.random.laplace(loc=0, scale=sensitivity / epsilon) for vec in vector]

# Выполнение проверки
vector = [1,2,3,4,5,6,7,8,9,10]
RES = [LaplaceMechVec(vector, 8, 1.0) for _ in range(200)]
for i, vec in enumerate(vector):
    s = [np.random.laplace(loc=vec, scale=8/1.0) for _ in range(200)]
```

К предоставленным в раздаточных материалах запросам Query1, Query2, Query3 для гарантии приватности полученных ответов, необходимо применить реализованный векторный механизм Гаусса для, а также механизм Лапласа, код которого указан выше, и написать функцию подсчета ошибок. Для удобства и наглядности последнюю требовалось изобразить в виде гистограммы, пример которой изображен на рисунке 23.

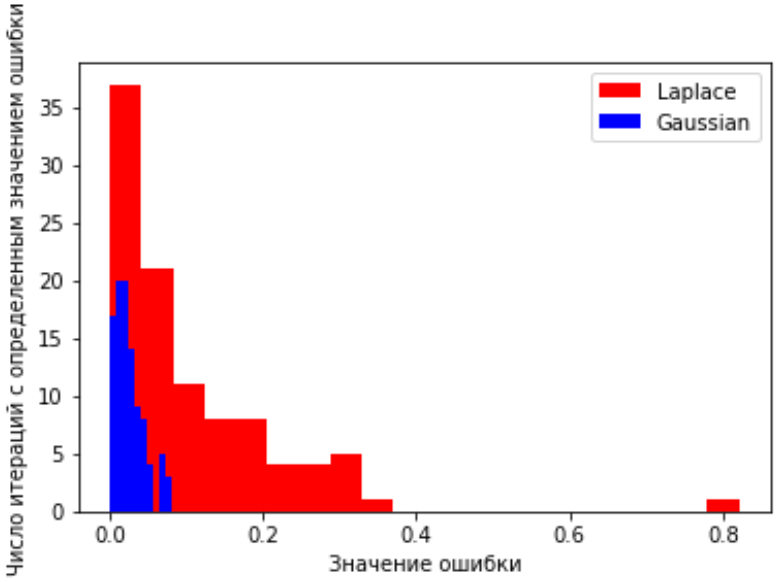

Рисунок 23 – График подсчета ошибок для результатов дифференциально защищенных механизмов Лапласа и Гаусса

Значения ошибок при использовании механизма Гаусса для обеспечения дифференциальной конфиденциальности нескольких запросов имеют меньшие значения, нежели в результате применения механизма Лапласа. Однако стоит учитывать, что механизм Гаусса работает лучше далеко не всегда. Он отлично подходит для обеспечения дифференциальной конфиденциальности при выполнении запросов к набору данных, где один человек оказывает влияние на множество величин. В противном случае, когда пользователь меняет только одно

значение, лучше использовать метод Лапласа. Также всегда стоит учитывать значение параметра δ, так как при больших значениях приватность данных может быть нарушена вовсе.

3.3.4 Работа 4 – Экспоненциальный механизм

Механизмы Лапласа и Гаусса, которые были рассмотрены в предыдущих лабораторных работах, сосредоточены на численных ответах и добавляют шум непосредственно к самому ответу. Если нужно получить точный ответ, который не содержит дополнительного шума, но при этом сохраняет дифференциальную конфиденциальность, используют экспоненциальный механизм. Последний позволяет выбрать "лучший" элемент из набора. Аналитик определяет, какой элемент является "лучшим", указывая функцию оценки, которая выводит оценку для каждого элемента в наборе данных. Например, предположим, что некая компания хочет установить цену на выпускаемое программное обеспечение. Чтобы сделать это, они проводят опрос потенциальных покупателей, спрашивая цену, которую те готовы заплатить. При этом людям обещают, что названные каждым из них, не будут разглашены, опубликуется лишь дифференциально конфиденциальное значение оптимальной стоимости. Однако, если руководствоваться механизмами Лапласа или Гаусса, к цене необходимо будет добавить шум для обеспечения дифференциальной приватности, что изменит значение цены и может оказаться таковым, что люди не захотят столько платить, а это приведет к нулевым продажам. Экспоненциальный механизм решает эту проблему. Учитывая параметр конфиденциальности Ɛ, набор результатов R и функцию оценки $u: D \times R \to \mathbb{R}$ с чувствительностью Δu, которая сопоставляет пары - (база данных, потенциальный результат) с оценкой, механизм выбирает один элемент $r \in R$ с вероятностью:

$$P[r \in R] = exp^{\frac{Ɛ * u(D,r)}{2\Delta u}}$$

Самое большое практическое различие между экспоненциальным механизмом и механизмами Лапласа и Гаусса заключается в том, что выход экспоненциального механизма всегда является членом множества R.

Алгоритм 1 – Экспоненциальный механизм для конечного множества

1. Вход:

a. набор данных D;

b. множество выходных значений R;

c. чувствительность Δu;

d. значение бюджета конфиденциальности ε.

2. Определить функцию оценки $u(d,r)$, которая будет вычислять «пользу» одного из выходных значений $r \in R$ относительно остальных, т. е. оценивать на сколько текущее значение лучше или больше остальных. Например, эта функция может иметь вид представленный ниже:

```
def u (d = adult['Marital Status'], r = 'Never-married'):
    return d.value_counts()[r]/np.max(d.value_counts())
```

3. Вычислить оценку: $u(d,r) \to rating$ для каждого элемента r из R.

4. Для каждого элемента $r \in R$ рассчитать вероятность на основе его оценки из пункта 2:

$$P[r \in R] = exp^{\frac{\varepsilon * u(D,r)}{2\Delta u}}$$

В качестве чувствительности можно принять значение $\Delta u = 1$.

5. Нормализовать вероятности, чтобы их сумма была равна 1.

6. Выбрать значение из R на основе вероятностей.

7. Выход: $r \in R$.

Экспоненциальный механизм интересен по нескольким причинам:

1. Механизм обеспечивает ε - дифференциальную конфиденциальность

2. Работает как для конечных, так и для бесконечных множеств R, однако, когда множество R бесконечно, может быть очень сложно построить практическую реализацию.

3. Механизм представляет собой "фундаментальный механизм" ε - дифференциальной конфиденциальности: все другие дифференциально частные механизмы могут быть определены в терминах экспоненциального с соответствующим определением функции оценки u.

Набор данных для анализа берется тот же, что был дан в лабораторной работе 1.

Помимо ранее указанных вариантов задания в этой работе, а именно запросы к данным, разделяются еще и по четности номера варианта:

– для четных вариантов необходимо определить семейный статус большинства в предоставляемом наборе данных запрос;

– для нечетных – образование большинства.

Затем по предоставленному алгоритму 1 необходимо реализовать экспоненциальный механизм для конечного множества, который будет вычислять вышеописанные запросы, и может иметь следующий вид:

```
def ExponentialMech(d, R, u, sensitivity = 1, epsilon = 1):
    # 1. Вычислить оценку для каждого элемента R
    rating = [u(d, r) for r in R]
    # 2. Рассчитать вероятность для каждого элемента на основе его оценки
    probabilities = [np.exp(epsilon * i / (2 * sensitivity)) for i in rating]
    # 3. Нормализовать вероятности, чтобы они в сумме равнялись 1
    probabilities = probabilities / np.linalg.norm(probabilities, ord=1)
    #4. Выбрать элемент из R на основе вероятностей
    return np.random.choice(R, 1, p=probabilities)[0]
```

Для вычисления точности, требуется выполнить 300 итераций экспоненциального механизма. Результат для нечетных вариантов может иметь вид, представленный на рисунке 24. Он показывает, что из 300-от итераций 230 выдали точный ответ - большинство людей в наборе закончили старшую школу, а еще 70 сделали ошибочные выводы, которые возникают из-за неудачного определения функции оценки:

```
def u (d = data['Education'], r):
    return d.value_counts()[r]/np.max(d.value_counts())
```

```
HS-grad         230
Some-college     34
Bachelors        20
Doctorate         3
11th              2
Prof-school       2
9th               2
10th              2
Masters           1
Assoc-acdm        1
12th              1
Preschool         1
5th-6th           1
dtype: int64
```

Рисунок 24 – Результат 300-от итераций экспоненциального механизма

При переопределении последней:

```
# Функция оценки
def EvaluationFunction(d,r):
    return d.value_counts()[r]/100
```

результат уже может иметь вид, представленный на рисунке 25, который дает нулевой процент ошибок.

```
HS-grad    300
dtype: int64
```

Рисунок 25 – Результат 300-от итераций экспоненциального механизма после переопределения функции оценки

На основании полученных знаний о различных механизмах обеспечения дифференциальной конфиденциальности, в данной лабораторной работе студенту предлагается проанализировать все ранее рассмотренные алгоритмы и заполнить сравнительную таблицу, которая в готовом виде представлена как таблица 5.

Таблица 5 - Сравнительная таблица механизмов обеспечения дифференциальной конфиденциальности

Параметры сравнения	Механизмы		
	Лапласа	Гаусса	Экспоненциальный
Уровень дифференциальной конфиденциальности	ε - дифференциальная конфиденциальность	(ε, δ) – дифференциальная конфиденциальность	ε - дифференциальная конфиденциальность

Таблица 5 - Сравнительная таблица механизмов обеспечения дифференциальной конфиденциальности (продолжение)

Параметры сравнения	Механизмы		
	Лапласа	Гаусса	Экспоненциальный
Как достигается конфиденциальность?	Добавляет шум к результату запроса в соответствии с распределением Лапласа. Количество шума определяется чувствительностью запроса.	Добавляет шум к результату запроса в соответствии с распределением Гаусса. Количество шума определяется чувствительностью запроса.	Результат определяется путем вычисления «пользы» параметра. В базе хранятся не сами значения, а оценка их «пользы».
Для каких запросов применим?	В основном численные запросы, где один человек влияет на один параметр в наборе данных.	В основном численные запросы, где один человек влияет на несколько параметров в наборе данных.	Запросы, где важна точность результата, или при выборе «лучшего» значения.
Чувствительность	L1	L1 или L2. Чувствительность L2 ниже L1, что позволяет добиться меньшего шума.	L1

***

Таким образом, данная глава описывает практическую часть, которая посвящена разработке лабораторных работ по обеспечению дифференциальной конфиденциальности. Здесь представлены причины выбора тем для разработанных работ, а также используемые при этом инструменты.

ЗАКЛЮЧЕНИЕ

С развитием технологий, обеспечивающих все более детальный и эффективный сбор информации, растет потребность в защите личной информации людей, предоставляющих свои персональные данные. Средства обработки данных должны гарантировать человеку приватность. Чтобы добиться этого необходимо надежное и математически строгое определение конфиденциальности, а также широкий класс алгоритмов, которые будут удовлетворять этому требованию.

В процессе исследования был произведен анализ существующих в настоящее время методов обеспечения конфиденциальности, в результате которого удалось выявить основные достоинства и недостатки каждого из них. На основании полученных результатов, был выбран метод, который на мой взгляд, является наиболее актуальным в современное время и будет актуален в ближайшем будущем. Этим методом является дифференциальная конфиденциальность. Конечно же, нельзя сказать, что последний не имеет минусов, однако метод широко применим известными компаниями, такими как Google, Uber, Microsoft и позволяет добиться конфиденциальности информации, обеспечивая при этом достаточную точность.

На основании этого выбора было решено разработать обучающий комплекс по обеспечению дифференциальной конфиденциальности, для чего сначала были проанализированы основные методы обеспечения дифференциальной приватности – механизм рандомизированного ответа, механизмы Лапласа и Гаусса и экспоненциальный механизм, а также сформулированы основные определения и свойства.

Результатом исследования стал разработанный обучающий комплекс, состоящий из четырех работ, который может позволить специалистам информационной безопасности оценивать меру конфиденциальности набора данных, понимать и применять основные методы обеспечения дифференциальной конфиденциальности, а также строить алгоритмы по защите приватности данных пользователей, чья личная информация содержится в

наборе данных. Работы были реализованы на языке Python в интерактивном веб-приложении Jupyter Notebook. Каждая лабораторная включают в себя цель, теоретические сведения, задания для работы, упражнения для понимания математического аппарата механизмов обеспечения дифференциальной конфиденциальности и для анализа полученных в ходе выполнения работы результатов, а также контрольные вопросы, ответы на которые можно найти в прилагаемой теоретической части, и содержание отчета. В состав раздаточного материала для студентов входят:

– набор данных, анализ которого производится в лабораторной работе;

– блокнот Jupyter Notebook с теоретической информацией, шаблонами кода и заданиями для работы;

– документ формата .docx с информацией, требующейся для выполнения лабораторной работы, упражнениями, контрольными вопросами и содержанием отчета.

Результаты работы могут быть использованы студентами по информационной безопасности, а также другими людьми, которые заинтересованы в теме обеспечения конфиденциальности данных.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Microsoft Find Cancer Clues in Search Queries. [Электронный ресурс]. URL: https://www.nytimes.com/2016/06/08/technology/online-searches-can-identify-cancer-victims-study-finds.html. – (дата обращения: 20.01.2021).

2. Федеральный закон №152-ФЗ «О персональных данных». [Электронный ресурс]. URL: https://rg.ru/2006/07/29/personaljnye-dannye-dok.html. – (дата обращения: 20.01.2021).

3. Приказ Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор) «Об утверждении требований и методов по обезличиванию персональных данных». [Электронный ресурс]. URL: https://rg.ru/2013/09/18/dannye-dok.html. – (дата обращения: 20.01.2021).

4. L. Sweeney k-anonymity: A model for protecting privacy//International Journal on Uncertainty, Fuzziness and Knowledge-based Systems. – 2002. –№. 5.– С. 557-570.

5. P. Samarati and L. Sweeney Generalizing data to provide anonymity when disclosing information // PODS. - 1998. - Т. 98. - №. 10.1145. - С. 275487.275508.

6. Liu K., Giannella C., Kargupta H. An attacker’s view of distance preserving maps for privacy preserving data mining //European Conference on Principles of Data Mining and Knowledge Discovery. – Springer, Berlin, Heidelberg, 2006. – С. 297-308.

7. Oliveira S. R., Za O. R. ıane,“Achieving privacy preservation when sharing data for clustering,” //Proc. Workshop Secure Data Manag. – 2004. – С. 67-82.

8. Chen K., Liu L. Privacy preserving data classification with rotation perturbation //Fifth IEEE International Conference on Data Mining (ICDM'05). – IEEE, 2005. – С. 4 pp.

9. Mivule K. Utilizing noise addition for data privacy, an overview //arXiv preprint arXiv:1309.3958. – 2013.

10. Chamikara M. A. P. et al. Efficient privacy preservation of big data for accurate data mining //Information Sciences. – 2020. – Т. 527. – С. 420-443.

11. Ishai Y. et al. Cryptography from Anonymity Ґ. – 2006.

12. Dwork C. Differential privacy: A survey of results //International conference on theory and applications of models of computation. – Springer, Berlin, Heidelberg, 2008. – С. 1-19.

13. Leonard M. Differential Privacy: Secure and Private AI, курс на Udacity. [Электронный ресурс]. URL: https://classroom.udacity.com/courses/ud185/lessons/3879a8dd-9393-4c7f-a615-5ac1bdb2a629/concepts/a5f2b5e4-e6c1-4b0d-9542-9c0b8102df68. – (дата обращения: 25.02.2021).

14. Dwork C. et al. The algorithmic foundations of differential privacy //Foundations and Trends in Theoretical Computer Science. – 2014. – Т. 9. – №. 3-4. – С. 211-407.

15. Wood A. et al. Differential privacy: A primer for a non-technical audience //Vand. J. Ent. & Tech. L. – 2018. – Т. 21. – С. 209.

16. Programming Differential Privacy. – 2020. [Электронный ресурс]. URL: https://uvm-plaid.github.io/programming-dp/notebooks/ch4.html. – (дата обращения: 14.04.2020).

17. Dwork C. et al. Our data, ourselves: Privacy via distributed noise generation //Annual International Conference on the Theory and Applications of Cryptographic Techniques. – Springer, Berlin, Heidelberg, 2006. – С. 486-503.

18. Cheruvu R. A High-level Introduction to Differential privacy. – 2018 [Электронный ресурс]. URL: https://demystifymachinelearning.wordpress.com/2018/11/20/intro-to-differential-privacy/. – (дата обращения: 14.03.2021).

19. Lundmark M., Dahlman C. J. Differential privacy and machine learning: Calculating sensitivity with generated data sets. – 2017.

20. Shaistha F., Query “Sensitivity” types and effects on Differential Privacy Mechanism. – 2020. [Электронный ресурс]. URL: https://becominghuman.ai/query-sensitivity-types-and-effects-on-differential-privacy-mechanism-c94fd14b9837/. – (дата обращения: 07.05.2021).

21. Shaistha F., Differential Privacy – Noise adding Mechanisms. – 2020. [Электронный ресурс]. URL: https://becominghuman.ai/differential-privacy-noise-adding-mechanisms-ede242dcbb2e. – (дата обращения: 24.05.2021).

22. Desfontaines D., The magic of Gaussian noise. – 2020. [Электронный ресурс]. URL: https://desfontain.es/privacy/gaussian-noise.html. – (дата обращения: 20.05.2021).

23. Jain P., Gyanchandani M., Khare N. Differential privacy: its technological prescriptive using big data //Journal of Big Data. – 2018. – Т. 5. – №. 1. – С. 15.

24. McSherry F., Talwar K. Mechanism design via differential privacy //48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07). – IEEE, 2007. – С. 94-103.

25. Microsoft Researching Blog. – 2017. [Электронный ресурс]. URL: https://www.microsoft.com/en-us/research/blog/project-privtree-blurring-location-privacy/. – (дата обращения: 27.05.2021).

26. Erlingsson Ú., Pihur V., Korolova A. Rappor: Randomized aggregatable privacy-preserving ordinal response //Proceedings of the 2014 ACM SIGSAC conference on computer and communications security. – 2014. – С. 1054-1067.

27. Goyal B., Agrawal S., Sohi B. S. Noise issues prevailing in various types of medical images //Biomedical & Pharmacology Journal. – 2018. – Т. 11. – №. 3. – С. 1227.

28. Documentation Jupyter. [Электронный ресурс]. URL: https://jupyter.org/documentation. – (дата обращения: 27.05.2021).

29. National Estimates of Drug-Related Emergency Department Visits. – 2014. [Электронный ресурс]. URL: https://www.samhsa.gov/data/report/national-estimates-drug-related-emergency-department-visits-2004-2011-all-visits. – (дата обращения: 17.05.2021).