Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

# BCI 2001 DATA PRIVACY

***A PROJECT REPORT ON***

# "IMPLEMENTATION OF HIPAA COMPLIANCES ON HEALTH CARE SYSTEM"

*Submitted in partial fulfillment of the requirements for the degree of*

**Bachelor of Technology**

*In*

**Computer Science and Engineering**

*by*


**KALYAN (20BCI0154)**

**AMAN SETH (20BKT0122)**

**PARAS CHAWLA (20BKT0116)**

**DHWAJ JAIN (20BCI0302)**

**TARUSH GUPTA (20BKT0024)**

*Under the guidance of*

**Prof. JASMIN T. JOSE**

School of Computer Science and Engineering

# DECLARATION

We hereby declare that the thesis entitled **"IMPLEMENTATION OF HIPAA COMPLIANCES ON HEALTH CARE SYSTEM"** submitted by our team, for the award of the degree of Bachelor of Technology in Computer Science to VIT, is a record of bonafide work carried out by our team under the guidance of Prof. Jasmin T. Jose.

We further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

**Place** : Vellore

**Date** : 18th November, 2022

**Signature of the Candidate**

kalyan (20BCI0154)

Aman Seth (20BKT0122)

Paras Chawla (20BKT0116)

Dhwaj Jain (20BCI0302)

Tarush Gupta (20BKT0024)

# CERTIFICATE

This is to certify that the thesis entitled **“IMPLEMENTATION OF HIPAA COMPLIANCES ON HEALTH CARE SYSTEM”** submitted by the team **Paras Chawla (20BKT0116), Aman Seth (20BKT0122), Dhwaj Jain (20BCI0302), Kalyan (20BCI0154), Tarush Gupta (20BKT0024),** SCOPE, VIT University, for the award of the degree of Bachelor of Technology in Computer Science, is a record of bonafide work carried out by them under my supervision during the period 20.07.2022 to 30.11.2022, as per the VIT code of academics and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The thesis fulfills the requirements and regulations of the University and in my opinion meets the necessary standards for submission.

**Place** : Vellore

**Date** : 18th November, 2022

**Signature of the Candidate**

kalyan (20BCI0154)

Aman Seth (20BKT0122)

Paras Chawla (20BKT0116)

Dhwaj Jain (20BCI0302)

Tarush Gupta (20BKT0024)

**Internal Examiner** **External Examiner**

Dr. Sathyaraj R

School of Computer Science and Engineering

# ACKNOWLEDGEMENT

We would like to express our gratitude and appreciation to all those who gave me the possibility to complete this project. Special thanks are due to my supervisor Prof. Jasmin T. Jose, whose help, stimulating suggestions and encouragement helped us in all time of fabrication process and in making this project. We also sincerely thanks for the time spent proofreading and correcting my many mistakes.

We would also like to acknowledge with much appreciation the crucial role of our HOD, Dr. Sathyaraj R, for providing us with this wonderful opportunity to work on this project. This project would not have been accomplished without their help and insights.

Many thanks go to the whole lecturer and supervisors who have given their full effort in guiding the team in achieving the goal as well as their encouragement to maintain our progress in track. My profound thanks go to all classmates, especially to my teammates for spending their time in helping and giving support whenever we need it in fabricating our project.

Finally, we wish to thank our parents for their support and encouragement throughout our study.

# ABSTRACT

Data privacy or information privacy is part of the data protection field, which is concerned with ensuring and complying with data protection laws. Generally, Data Privacy or Information Privacy encompasses three components: the right of an individual to control their personal information, the policies for processing, collecting, storing, and sharing that information, and compliance with data protection laws.

Health Insurance Portability and Accountability Act (HIPAA) is one of the laws that demands confidentiality and privacy protection of healthcare data of individuals. In a stored database of a patient in a hospital or a clinic, we can develop a conservational and analytical method to keep the medical records of the patients in a well-preserved and adequate environment. This will ensure privacy of data and maintain its utility as well.

# INTRODUCTION

The purpose of HIPAA is to ensure that patient or customer Private Health Information (PHI) remains private. In order to protect healthcare data, HIPAA requires that such measures be taken by businesses, companies and healthcare organizations. HIPAA compliance might seem daunting, but a step-by-step approach can get you there. Our objective is to collect data sets related to the health department and apply HIPAA security rules to protect the data.

Most medical and health data not covered by HIPAA are controlled by third party data brokers and Internet companies. These companies combine this data with a wide range of personal information about consumer daily activities, transactions, movements, and demographics. The combined data are used for predictive profiling of individual health status, and often sold for advertising and other purposes. The rapid expansion of medical and health data outside of HIPAA protection is encroaching on privacy and doctor-patient relationship. This becomes even more of a concern when the medical departments involved are ones like psychiatry.

In our project we have taken a database that contains patient data in an un-anonymized form and then we apply various data protection algorithms so that privacy is achieved and its utility is also not compromised.

## LITERATURE REVIEW

| S.No. | Year of Publication | Authors and Title | Contribution | Limitations |
|---|---|---|---|---|
| **1.** | 2009 | Mbonihankuye, S., Nkunzimana, A., & Ndagijimana, A. (2019). Healthcare data security technology: HIPAA compliance. *Wireless communications and mobile computing* | The HIS is a system that aims to provide internal and external communication among healthcare providers.<br><br>HIS provides a common source of information about a patient's health history. | In medical testing, some binary classifications may find a false positive which results in some errors in data reporting when the test result improperly indicates presence of a condition such as a disease.<br><br>It sometimes contains false-negative error which improperly indicates the no presence of data condition and information security. |
| **2.** | 2014 | Glenn, T., & Monteith, S. Privacy in the digital world: medical and health data outside of HIPAA protections. *Current psychiatry reports* | This review will focus on the medical and health data that are increasingly being collected outside of HIPAA protections.<br><br>Medical and health data outside of HIPAA can be volunteered by consumers directly, observed by corporations recording consumer actions, and inferred by calculated models | Digital data does not reside where it was generated. Data moves and is serviced by many corporations and devices, including Internet service providers, etc.<br><br>About one-fourth of all digital data are original information, while the remaining three-fourths are duplications such as email attachments and backup copies |
| **3.** | 2014 | Nxumalo, Z. C., Tarwireyi, P., & Adigun, M. O. Towards privacy with | The solution presented in this paper allows GUISET | Most applications that are deployed in the GUISET context process |

| | | | | |
|---|---|---|---|---|
| | | tokenization as a service. In *2014 IEEE 6th International Conference on Adaptive Science & Technology* | applications to process sensitive user information in a safe manner.<br><br>It allows the removal of data from the processing environment thus decreasing the impacts of data breaches. The solution is easy to implement and is not resource intensive.<br><br>It also helps organizations to achieve PCI compliance at a low cost. | sensitive data like credit card information and protected health information which calls for the need to implement a measure to ensure privacy. |
| **4.** | 2016 | Sajid, A., & Abbas, H. Data privacy in cloud-assisted healthcare systems: state of the art and future challenges. *Journal of medical systems*, *40*(6), 1-16. | It helped to find precise answers to our defined research questions. The patient's data privacy concerns were identified and their corresponding mechanisms were also found from the selected literature.<br><br>The review revealed the fact that, the most applied technique to address the patient's data privacy concerns in healthcare cloud are IBE, ABE and its variants. | One of the top security concerns to the healthcare systems is to provide healthcare data privacy.<br><br>If the patients involved in the healthcare cloud systems are not ensured about their data's privacy, they will refuse to utilize these beneficial systems. |
| **5.** | 2019 | Paul, S., Joy, J. I., Sarker, S., Ahmed, S., & Das, A. K. Fake news detection in social media using blockchain. | Despite having some limitations, the proposed method will be undoubtedly helpful for | Using the Ethereum blockchain, it's difficult to detect the news based on politics and religion. |

| | | | | |
|---|---|---|---|---|
| | | | detecting fake news in social media as spreading fake news via social media which is a huge issue.<br><br>This news misguides people just to achieve more page views to earn extra money dishonestly. | For its veridical verification system, journals and news portals have to face job risk as it drives them to a competition of obtaining ratings. |

# PROPOSED METHODOLOGY

Our dataset comprises 11 fields. We can categorize them into different attribute types and apply the privacy protection methods accordingly: -

## EXPLICIT IDENTIFIERS (EI's)

Attributes that identify a customer/record owner directly. These include attributes like social security number (SSN), insurance ID, name, etc.

**i) patient_id:** It is 11 characters long and is unique for every patient. It can be used to directly identify a patient. We have applied a custom tokenization algorithm on it to prevent identity disclosure. This involves passing the string through SHA-256 first and then through Keccak. Both are very secure hashing algorithms and are in widespread use for many different applications. They will generate a 56-character long string which we will truncate down to 11 characters again using the random function in python. This will generate a completely unique token for all patient ids and in this way the possible adversary won't have any knowledge of what method was used to generate the token.

**ii) patient_last_name:** This can also be used to uniquely identify patients. To avoid this, we have completely suppressed the last name.

## QUASI IDENTIFIERS (QI's)

Attributes that include geographic and demographic information, phone numbers, and e-mail IDs. Quasi identifiers are also defined as those attributes that are publicly available, for example, a voters database.

**iii) patient_gender:** It can have values like 'M' to identify as male or 'F' to identify as female. We have not anonymised it as it doesn't cause direct loss of privacy and also doctors can know about the gender of their patient to treat them better.

**iv) patient_age:** This tells the age of a patient. This field combined with other quasi-identifiers increases the chance of background knowledge attack and rediscovery. We have generalized it by creating intervals like <15, 15-30, 30-50 and >50.

**v) patient_first_initial:** As it suggests, it stores the first initial of the patient. Since we have completely suppressed the last name, it need not be anonymised, as this alone cannot be used to identify patients.

**vi) patient_race:** This tells the race of a patient. Although this information can be used in identifying a person, we will not anonymise it as it can cause a high loss in utility if done so. This is because doctors sometimes gain valuable information for a diagnosis, based on the individual's race.

## SENSITIVE DATA (SD)

Attributes that contain confidential data information about the record owner, such as health issues, financial status, and salary, which cannot be compromised at any cost.

**vii) patient_sat_score:** Patient satisfaction score is a rating which a patient can give to the hospital based on their quality of service. We have applied randomization on this field by using laplace truncated method. This method will randomize the values till one decimal place. The purpose behind applying this particular method is that it retains the statistical utility of the numerical data while randomizing the values for each patient.

**viii) patient_admin_flag:** This tells us whether a patient is admitted in the hospital or not. Although this data is sensitive, we will not anonymize it since it would lead to a high loss in utility of data.

**ix) department_referral:** It indicates which medical department the patient was referred to. Again, anonymising this will lead to high loss in utility so we will not anonymize it. A patient's privacy protection will be mainly achieved through anonymization of explicit and quasi-identifiers.

## NON SENSITIVE DATA (NSD)

Data that is not sensitive for the given context.

**x) date:** Reveals date of visit of patient. Since this is non-sensitive data, we don't need to do anything about it.

**xi) patient_waittime:** Tells how long the patient had to wait before seeing the doctor. Hospitals can use this information to improve their services. But it does not reveal any information about the patient and hence it can be left as it is.

## IMPLEMENTATION

**CODE:**

```
!pip install pycryptodome

!pip install diffprivlib

import random;

from Crypto.Hash import keccak

import pandas as pd;

import hashlib

healthcare=pd.read_csv("/content/Hospital ER.csv");


from numbers import Real


import numpy as np


from        diffprivlib.mechanisms.base        import        DPMechanism,
TruncationAndFoldingMixin

from diffprivlib.utils import copy_docstring



class Laplace(DPMechanism):


   def __init__(self, *, epsilon, delta=0.0, sensitivity):

       super().__init__(epsilon=epsilon, delta=delta)

       self.sensitivity = self._check_sensitivity(sensitivity)

       self._scale = None


   @classmethod

   def _check_sensitivity(cls, sensitivity):
```

```
        if not isinstance(sensitivity, Real):
            raise TypeError("Sensitivity must be numeric")

        if sensitivity < 0:
            raise ValueError("Sensitivity must be non-negative")

        return float(sensitivity)

    def _check_all(self, value):
        super()._check_all(value)
        self._check_sensitivity(self.sensitivity)

        if not isinstance(value, Real):
            raise TypeError("Value to be randomised must be a number")

        return True

    def bias(self, value):

        return 0.0

    def variance(self, value):

        self._check_all(0)

        return 2 * (self.sensitivity / (self.epsilon - np.log(1 -
self.delta))) ** 2

    @staticmethod
```

```
    def _laplace_sampler(unif1, unif2, unif3, unif4):
        return np.log(1 - unif1) * np.cos(np.pi * unif2) + np.log(1 - unif3) * np.cos(np.pi * unif4)

    def randomise(self, value):

        self._check_all(value)

        scale = self.sensitivity / (self.epsilon - np.log(1 - self.delta))
        standard_laplace = self._laplace_sampler(self._rng.random(), self._rng.random(), self._rng.random(),
                                                 self._rng.random())

        return value - scale * standard_laplace


class LaplaceTruncated(Laplace, TruncationAndFoldingMixin):

    def __init__(self, *, epsilon, delta=0.0, sensitivity, lower, upper):
        super().__init__(epsilon=epsilon, delta=delta, sensitivity=sensitivity)
        TruncationAndFoldingMixin.__init__(self, lower=lower, upper=upper)

    @copy_docstring(Laplace.bias)
    def bias(self, value):
        self._check_all(value)

        shape = self.sensitivity / self.epsilon
```

```
        return shape / 2 * (np.exp((self.lower - value) / shape) - np.exp((value - self.upper) / shape))

    @copy_docstring(Laplace.variance)
    def variance(self, value):
        self._check_all(value)

        shape = self.sensitivity / self.epsilon

        variance = value ** 2 + shape * (self.lower * np.exp((self.lower - value) / shape)
                                         - self.upper * np.exp((value - self.upper) / shape))
        variance += (shape ** 2) * (2 - np.exp((self.lower - value) / shape)
                                    - np.exp((value - self.upper) / shape))

        variance -= (self.bias(value) + value) ** 2

        return variance

    def _check_all(self, value):
        Laplace._check_all(self, value)
        TruncationAndFoldingMixin._check_all(self, value)

        return True

    @copy_docstring(Laplace.randomise)
    def randomise(self, value):
        self._check_all(value)
```

```
        noisy_value = super().randomise(value)
        return self._truncate(noisy_value)


class LaplaceFolded(Laplace, TruncationAndFoldingMixin):

    def __init__(self, *, epsilon, delta=0.0, sensitivity, lower, upper,
random_state=None):
        super().__init__(epsilon=epsilon,                    delta=delta,
sensitivity=sensitivity, random_state=random_state)
        TruncationAndFoldingMixin.__init__(self,             lower=lower,
upper=upper)

    @copy_docstring(Laplace.bias)
    def bias(self, value):
        self._check_all(value)

        shape = self.sensitivity / self.epsilon

        bias = shape * (np.exp((self.lower + self.upper - 2 * value) /
shape) - 1)
        bias /= np.exp((self.lower - value) / shape) + np.exp((self.upper
- value) / shape)

        return bias

    @copy_docstring(DPMechanism.variance)
    def variance(self, value):
        raise NotImplementedError
```

```
    def _check_all(self, value):
        super()._check_all(value)
        TruncationAndFoldingMixin._check_all(self, value)

        return True

    @copy_docstring(Laplace.randomise)
    def randomise(self, value):
        self._check_all(value)

        noisy_value = super().randomise(value)
        return self._fold(noisy_value)


class LaplaceBoundedDomain(LaplaceTruncated):

    def _find_scale(self):
        eps = self.epsilon
        delta = self.delta
        diam = self.upper - self.lower
        delta_q = self.sensitivity

        def _delta_c(shape):
            if shape == 0:
                return 2.0
            return  (2  -  np.exp(-  delta_q  /  shape)  -  np.exp(-  (diam  -
delta_q) / shape)) / (1 - np.exp(- diam / shape))

        def _f(shape):
```

```
            return delta_q / (eps - np.log(_delta_c(shape)) - np.log(1 -
delta))

        left = delta_q / (eps - np.log(1 - delta))
        right = _f(left)
        old_interval_size = (right - left) * 2

        while old_interval_size > right - left:
            old_interval_size = right - left
            middle = (right + left) / 2

            if _f(middle) >= middle:
                left = middle
            if _f(middle) <= middle:
                right = middle

        return (right + left) / 2

    def effective_epsilon(self):

        if self._scale is None:
            self._scale = self._find_scale()

        if self.delta > 0.0:
            return None

        return self.sensitivity / self._scale

    @copy_docstring(Laplace.bias)
```

```python
    def bias(self, value):
        self._check_all(value)

        if self._scale is None:
            self._scale = self._find_scale()

        bias = (self._scale - self.lower + value) / 2 * np.exp((self.lower - value) / self._scale) \
            - (self._scale + self.upper - value) / 2 * np.exp((value - self.upper) / self._scale)
        bias /= 1 - np.exp((self.lower - value) / self._scale) / 2 \
            - np.exp((value - self.upper) / self._scale) / 2

        return bias

    @copy_docstring(Laplace.variance)
    def variance(self, value):
        self._check_all(value)

        if self._scale is None:
            self._scale = self._find_scale()

        variance = value**2
        variance -= (np.exp((self.lower - value) / self._scale) * (self.lower ** 2)
                     + np.exp((value - self.upper) / self._scale) * (self.upper ** 2)) / 2
        variance += self._scale * (self.lower * np.exp((self.lower - value) / self._scale)
                                   - self.upper * np.exp((value - self.upper) / self._scale))
```

```
        variance += (self._scale ** 2) * (2 - np.exp((self.lower - value) / self._scale)
                                          - np.exp((value - self.upper) / self._scale))
        variance /= 1 - (np.exp(-(value - self.lower) / self._scale)
                         + np.exp(-(self.upper - value) / self._scale)) / 2

        variance -= (self.bias(value) + value) ** 2

        return variance

    @copy_docstring(Laplace.randomise)
    def randomise(self, value):
        self._check_all(value)

        if self._scale is None:
            self._scale = self._find_scale()

        value = max(min(value, self.upper), self.lower)
        if np.isnan(value):
            return float("nan")

        samples = 1

        while True:
            try:
                unif = self._rng.random(4 * samples)
            except TypeError:  # rng is secrets.SystemRandom
                unif = [self._rng.random() for _ in range(4 * samples)]
```

```
            noisy        =        value        +        self._scale        *
self._laplace_sampler(*np.array(unif).reshape(4, -1))


            if ((noisy >= self.lower) & (noisy <= self.upper)).any():
                idx  =  np.argmax((noisy  >=  self.lower)  &  (noisy  <=
self.upper))
                return noisy[idx]
            samples = min(100000, samples * 2)



class LaplaceBoundedNoise(Laplace):


    def __init__(self, *, epsilon, delta, sensitivity, random_state=None):
        super().__init__(epsilon=epsilon,                    delta=delta,
sensitivity=sensitivity, random_state=random_state)
        self._noise_bound = None


    @classmethod
    def _check_epsilon_delta(cls, epsilon, delta):
        if epsilon == 0:
            raise ValueError("Epsilon must be strictly positive. For zero
epsilon, use :class:`.Uniform`.")


        if isinstance(delta, Real) and not 0 < delta < 0.5:
            raise  ValueError("Delta  must  be  strictly  in  the  interval
(0,0.5). For zero delta, use :class:`.Laplace`.")


        return super()._check_epsilon_delta(epsilon, delta)


    @copy_docstring(Laplace.bias)
    def bias(self, value):
```

```
        return 0.0

    @copy_docstring(DPMechanism.variance)
    def variance(self, value):
        raise NotImplementedError

    @copy_docstring(Laplace.randomise)
    def randomise(self, value):
        self._check_all(value)

        if self._scale is None or self._noise_bound is None:
            self._scale = self.sensitivity / self.epsilon
            self._noise_bound = 0 if self._scale == 0 else \
                self._scale * np.log(1 + (np.exp(self.epsilon) - 1) / 2 / self.delta)

        if np.isnan(value):
            return float("nan")

        samples = 1

        while True:
            try:
                unif = self._rng.random(4 * samples)
            except TypeError:  # rng is secrets.SystemRandom
                unif = [self._rng.random() for _ in range(4 * samples)]
            noisy = self._scale * self._laplace_sampler(*np.array(unif).reshape(4, -1))
```

```
            if ((noisy >= - self._noise_bound) & (noisy <= self._noise_bound)).any():
                idx = np.argmax((noisy >= - self._noise_bound) & (noisy <= self._noise_bound))
                return value + noisy[idx]
            samples = min(100000, samples * 2)



last_names=healthcare["patient_last_name"];
ids=healthcare["patient_id"];
sat=healthcare["patient_sat_score"];
sensitivity=3
epsilon=0.3
mechanism = LaplaceTruncated(epsilon=epsilon,delta=0.0,sensitivity=sensitivity,lower=0,upper=10)
new_sat=[];
for x in sat:
 new_sat.append(mechanism.randomise(x));
healthcare["patient_sat_score"]=new_sat;
age=healthcare["patient_age"];
new_age=[]
for x in age:
 if x<=15:
   new_age.append("<15");
 elif x<=30:
   new_age.append("15-30");
 elif x<=50:
   new_age.append("30-50");
 else:
```

```
   new_age.append(">50");
 healthcare["patient_age"]=new_age



def hash_unicode(a_string):
   sha_out= hashlib.sha256(a_string.encode('utf-8')).hexdigest()
   keccak_hash3 = keccak.new(digest_bits=224)
   keccak_hash3.update(sha_out.encode("utf-8"))
   hex=keccak_hash3.hexdigest();
   out="";
   for x in range(11):
     num=random.randint(0,55);
     out += hex[num];
   return out;
def supperession(a_string):
 return "XXXXXXXXX"
healthcare['patient_id']=healthcare['patient_id'].apply(hash_unicode);
healthcare['patient_last_name']=healthcare['patient_last_name'].apply(s
upperession)
healthcare.head()
from pathlib import Path
filepath = Path('content/subfolder/out.csv')
filepath.parent.mkdir(parents=True, exist_ok=True)
healthcare.to_csv(filepath)
```

# INPUT

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | date | patient_id | patient_gender | patient_age | patient_sat_score | patient_first_inital | patient_last_name | patient_race | patient_admin_flag | patient_waittime | department_referral | | |
| 2 | 2020-03-20 08:47:01 | 145-39-5406 | M | 69 | 10 | H | Glasspool | White | false | 39 | None | | |
| 3 | 2020-06-15 11:29:36 | 316-34-3057 | M | 4 | | X | Methuen | Native American/Alaska Native | true | 27 | None | | |
| 4 | 2020-06-20 09:13:13 | 897-46-3852 | F | 56 | 9 | P | Schubuser | African American | true | 55 | General Practice | | |
| 5 | 2020-02-04 22:34:29 | 358-31-9711 | F | 24 | 8 | U | Titcombe | Native American/Alaska Native | true | 31 | General Practice | | |
| 6 | 2020-09-04 17:48:27 | 289-26-0537 | M | 5 | | Y | Gionettitti | African American | false | 10 | Orthopedics | | |
| 7 | 2019-04-20 00:13:05 | 255-51-2877 | M | 58 | | H | Buff | Asian | false | 59 | None | | |
| 8 | 2019-08-23 08:26:21 | 465-97-0990 | F | 68 | | F | Perrat | White | true | 43 | None | | |
| 9 | 2019-07-29 16:57:15 | 157-31-7520 | F | 47 | | K | Gwillim | Two or More Races | true | 23 | None | | |
| 10 | 2020-02-19 06:54:39 | 432-34-5614 | F | 79 | 1 | E | Dewhirst | White | false | 42 | None | | |
| 11 | 2020-10-11 05:25:17 | 609-17-8678 | M | 62 | | M | Crebo | African American | false | 51 | None | | |
| 12 | 2020-07-26 01:45:45 | 497-14-6812 | F | 73 | | Q | Churchard | White | true | 34 | Gastroenterology | | |
| 13 | 2020-03-10 22:02:15 | 393-38-9502 | F | 16 | | N | Corpes | White | false | 39 | Orthopedics | | |
| 14 | 2019-11-12 16:00:12 | 288-05-6370 | F | 16 | | R | Brixey | Native American/Alaska Native | true | 53 | General Practice | | |
| 15 | 2019-06-25 09:40:39 | 784-54-9931 | M | 46 | | M | Goudie | Pacific Islander | false | 45 | None | | |
| 16 | 2019-05-04 13:16:12 | 662-21-6522 | M | 69 | | G | Stanlack | White | true | 49 | None | | |
| 17 | 2019-09-19 01:53:21 | 628-73-1801 | M | 37 | | C | McMurty | Declined to Identify | true | 57 | None | | |
| 18 | 2020-05-25 22:11:20 | 370-19-2271 | F | 50 | | I | Scothorn | Asian | false | 35 | General Practice | | |
| 19 | 2019-06-25 18:59:56 | 458-98-8860 | M | 37 | | J | Helgass | Declined to Identify | false | 55 | None | | |
| 20 | 2019-09-04 16:15:52 | 728-31-2493 | F | 70 | | W | Chittock | Asian | true | 50 | Physiotherapy | | |
| 21 | 2019-11-16 23:46:29 | 823-34-5523 | M | 55 | 2 | F | Prendergast | Asian | true | 40 | None | | |
| 22 | 2019-06-30 05:22:02 | 621-70-7472 | F | 63 | | T | Bissiker | Native American/Alaska Native | true | 25 | None | | |
| 23 | 2019-05-22 16:48:52 | 344-36-7156 | F | 44 | 2 | M | Mandell | Asian | false | 51 | None | | |
| 24 | 2019-11-17 07:24:09 | 455-21-3671 | F | 11 | 4 | D | Coste | Declined to Identify | false | 30 | None | | |
| 25 | 2020-01-26 05:58:48 | 259-10-0339 | M | 8 | | Q | Dodridge | Asian | false | 16 | General Practice | | |
| 26 | 2019-05-24 14:42:43 | 720-54-2625 | F | 4 | | C | Pavie | Native American/Alaska Native | true | 23 | None | | |
| 27 | 2019-04-12 21:02:24 | 661-92-7059 | M | 42 | 0 | Z | Sleightholm | African American | false | 51 | None | | |
| 28 | 2020-09-04 02:30:09 | 598-53-3927 | F | 68 | | S | Noads | African American | false | 58 | None | | |
| 29 | 2019-12-16 13:02:55 | 715-74-5338 | M | 22 | | J | Filkin | Pacific Islander | false | 25 | General Practice | | |
| 30 | 2020-03-11 17:06:20 | 669-74-2146 | F | 58 | | C | Bilby | Declined to Identify | false | 55 | Orthopedics | | |
| 31 | 2020-07-07 14:58:01 | 693-38-2084 | F | 72 | | H | Kehoe | African American | true | 37 | None | | |
| 32 | 2020-04-02 13:19:49 | 548-93-9953 | F | 73 | | S | O'Neill | Two or More Races | false | 46 | Neurology | | |
| 33 | 2019-06-30 22:38:01 | 334-76-4005 | M | 35 | | J | Sibbit | White | true | 40 | Orthopedics | | |
| 34 | 2020-08-05 15:17:49 | 846-66-7490 | M | 48 | | T | Clissett | White | false | 53 | None | | |
| 35 | 2020-08-05 01:33:44 | 278-49-6531 | F | 74 | 0 | L | Vannacci | White | false | 20 | Physiotherapy | | |

# OUTPUT

| | date | patient_id | patient_gender | patient_age | patient_sat_score | patient_first_inital | patient_last_name | patient_race | patient_admin_flag | patient_waittime | department_referral |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2020-03-20 08:47:01 | abf3eb66355 | M | >50 | 10 | H | XXXXXXXXX | White | False | 39 | None |
| 1 | 2020-06-15 11:29:36 | a8b9bb0966f | M | <15 | | X | XXXXXXXXX | Native American/Alaska Native | True | 27 | None |
| 2 | 2020-06-20 09:13:13 | 659e5ee548b | F | >50 | 2.71384609790195 | P | XXXXXXXXX | African American | True | 55 | General Practice |
| 3 | 2020-02-04 22:34:29 | 2906f7ff0f5 | F | 15-30 | 0 | U | XXXXXXXXX | Native American/Alaska Native | True | 31 | General Practice |
| 4 | 2020-09-04 17:48:27 | dada5ab8a96 | M | <15 | | Y | XXXXXXXXX | African American | False | 10 | Orthopedics |
| 5 | 2019-04-20 00:13:05 | 5634bb64bbd | M | >50 | | H | XXXXXXXXX | Asian | False | 59 | None |
| 6 | 2019-08-23 08:26:21 | 9cdbadb7bab | F | >50 | | F | XXXXXXXXX | White | True | 43 | None |
| 7 | 2019-07-29 16:57:15 | 32780e57273 | F | 30-50 | | K | XXXXXXXXX | Two or More Races | True | 23 | None |
| 8 | 2020-02-19 06:54:39 | 84823182241 | F | >50 | 0 | E | XXXXXXXXX | White | False | 42 | None |
| 9 | 2020-10-11 05:25:17 | 1bb203da81e | M | >50 | | M | XXXXXXXXX | African American | False | 51 | None |
| 10 | 2020-07-26 01:45:45 | 3ba9ad5eafd | F | >50 | | Q | XXXXXXXXX | White | True | 34 | Gastroenterology |
| 11 | 2020-03-10 22:02:15 | c9487f42e19 | F | 15-30 | | N | XXXXXXXXX | White | False | 39 | Orthopedics |
| 12 | 2019-11-12 16:00:12 | 42bc492138d | F | 15-30 | | R | XXXXXXXXX | Native American/Alaska Native | True | 53 | General Practice |
| 13 | 2019-06-25 09:40:39 | ffd36cb3dd9 | M | 30-50 | | M | XXXXXXXXX | Pacific Islander | False | 45 | None |
| 14 | 2019-05-04 13:16:12 | 5cd2dbb969f | M | >50 | | G | XXXXXXXXX | White | True | 49 | None |
| 15 | 2019-09-19 01:53:21 | aa2d09e3ada | M | 30-50 | | C | XXXXXXXXX | Declined to Identify | True | 57 | None |
| 16 | 2020-05-25 22:11:20 | a8e26555b5c | F | 30-50 | | I | XXXXXXXXX | Asian | False | 35 | General Practice |
| 17 | 2019-06-25 18:59:56 | d91775998c9 | M | 30-50 | | J | XXXXXXXXX | Declined to Identify | False | 55 | None |
| 18 | 2019-09-04 16:15:52 | 7f29451e6f6 | F | >50 | | W | XXXXXXXXX | Asian | True | 50 | Physiotherapy |
| 19 | 2019-11-16 23:46:29 | 59c515f5373 | M | >50 | 0 | F | XXXXXXXXX | Asian | True | 40 | None |
| 20 | 2019-06-30 05:22:02 | 5669858f0e1 | F | >50 | | T | XXXXXXXXX | Native American/Alaska Native | True | 25 | None |
| 21 | 2019-05-22 16:48:52 | c24b239dc4a | F | 30-50 | 10 | M | XXXXXXXXX | Asian | False | 51 | None |
| 22 | 2019-11-17 07:24:09 | c5d9011c8df | F | <15 | 0 | D | XXXXXXXXX | Declined to Identify | False | 30 | None |
| 23 | 2020-01-26 05:58:48 | 790028dab3c | M | <15 | | Q | XXXXXXXXX | Asian | False | 16 | General Practice |
| 24 | 2019-05-24 14:42:43 | 7fe3ed7db3a | F | <15 | | C | XXXXXXXXX | Native American/Alaska Native | True | 23 | None |
| 25 | 2019-04-12 21:02:24 | 9484afeb3e8 | M | 30-50 | 0 | Z | XXXXXXXXX | African American | False | 51 | None |
| 26 | 2020-09-04 02:30:09 | 3c42550aef3 | F | >50 | | S | XXXXXXXXX | African American | False | 58 | None |
| 27 | 2019-12-16 13:02:55 | 1a6b674d952 | M | 15-30 | | J | XXXXXXXXX | Pacific Islander | False | 25 | General Practice |
| 28 | 2020-03-11 17:06:20 | 6276219669 | F | >50 | | C | XXXXXXXXX | Declined to Identify | False | 55 | Orthopedics |
| 29 | 2020-07-07 14:58:01 | 7da5a453edb | F | >50 | | H | XXXXXXXXX | African American | True | 37 | None |
| 30 | 2020-04-02 13:19:49 | a8ba1979176 | F | >50 | | S | XXXXXXXXX | Two or More Races | False | 46 | Neurology |
| 31 | 2019-06-30 22:38:01 | 1344b74a3e5 | M | 30-50 | | J | XXXXXXXXX | White | True | 40 | Orthopedics |
| 32 | 2020-08-05 15:17:49 | ce1a3a6dc8d | M | 30-50 | | T | XXXXXXXXX | White | False | 53 | None |
| 33 | 2020-08-05 01:33:44 | 962d0b66c71 | F | >50 | 8.4239174041324 | L | XXXXXXXXX | White | False | 20 | Physiotherapy |

# PROJECT DEMONSTRATION VIDEO

[https://drive.google.com/drive/folders/1eOs3Ie13tvcO8YCRa0vVPX0e985TziVK?usp=sharing](https://drive.google.com/drive/folders/1eOs3Ie13tvcO8YCRa0vVPX0e985TziVK?usp=sharing)

## CONCLUSION/FUTURE WORK

Large quantities of health data are being created outside of HIPAA protection, primarily by consumers. Most of the data generated by consumers are controlled by data brokers and Internet companies that have no involvement in patient care and no training in medical ethics. Data brokers are combining health data with other consumer data to make health related profiles, which may increasingly be used to identify individual health status. The results of the predictive profiles may have adverse impact regardless of accuracy. As knowledge of data brokers becomes more widespread, more patients may avoid healthcare or withhold data from physicians due to privacy concerns, which may have especially serious consequences in psychiatry.

The far-reaching problems relating to the use and protection of medical and health data outside of HIPAA need to be addressed by broad collaborations of medical, legal, consumer, and technical expertise. In the interim, measures to increase awareness of the growth of medical and health data outside of HIPAA protection are needed for both clinicians and patients.

# REFERENCES


**[1]** Mbonihankuye, S., Nkunzimana, A., & Ndagijimana, A. (2019). Healthcare data security technology: HIPAA compliance. *Wireless communications and mobile computing*, *2019*.

**[2]** Glenn, T., & Monteith, S. (2014). Privacy in the digital world: medical and health data outside of HIPAA protections. *Current psychiatry reports*, *16*(11), 1-11.

**[3]** Nxumalo, Z. C., Tarwireyi, P., & Adigun, M. O. (2014, October). Towards privacy with tokenization as a service. In *2014 IEEE 6th International Conference on Adaptive Science & Technology (ICAST)* (pp. 1-6). IEEE.

**[4]** Sajid, A., & Abbas, H. (2016). Data privacy in cloud-assisted healthcare systems: state of the art and future challenges. *Journal of medical systems*, *40*(6), 1-16.

**[5]** Paul, S., Joy, J. I., Sarker, S., Ahmed, S., & Das, A. K. (2019, June). Fake news detection in social media using blockchain. In *2019 7th International Conference on Smart Computing & Communications (ICSCC)* (pp. 1-5). IEEE.