International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : 10.32628/CSEIT206240

Predictive Disease Modelling of Differentially Private Health Data

Nitha V R
Department of Computer Science, Sree Narayana College, Cherthala, Alapuzha, Kerala, India

ABSTRACT

There is a range of fitness bands, smart watches and accessories that help you to track your steps, heart rate, weight, blood pressure, pulse rate, water intake, number of calories burned, location etc. Fitbit is such a company which has products in two main categories: smart watches and fitness trackers. Fitbit is a wearable wireless device that can recognize your running rate, biking speed and even swimming rate thereby monitoring your progress day by day or week by week and adjust your goals accordingly. If your Fitbit has GPS facility, you can see a map of your route, pace and elevation also. Fitbit use an algorithm to look for motions that indicate walking or running rate of a person. With the help of Fitbit dashboard, you can track your overall fitness and health.

Keywords : Fitbit, GPS, Differential Privacy

I. INTRODUCTION

SECURITY THREAT TO DATA PRIVACY

Many companies are planning to buy the health and fitness data of individuals either for data analysis, research or for privacy breach. Google, the tech giant started a project called 'Project Nightingale' that gathers personal Health Data on millions of Americans. Google has decided to buy the health tracking Device Company, Fitbit for $2.1 billion has raised concern over how user's data will be used.

DIFFERENTIAL PRIVACY

In 21st century, we heard many big data breaches that made people worry about the data privacy. In addition to this, most of the machine learning come from learning techniques which require large amount of training data. Along with this, research institutions often use and share data containing sensitive or confidential information about individuals. There is a chance of adverse consequences for a data's private information if the data is improperly disclosed. This led to the development of a formal privacy model for solving data privacy issue.

Many organizations are applying different privacy to protect sensitive information. Hence, Differential Privacy (DP) permits companies to access large number of sensitive data for research and business without any privacy breach. DP is a strong mathematical definition of privacy in context of statistical and machine learning analysis.

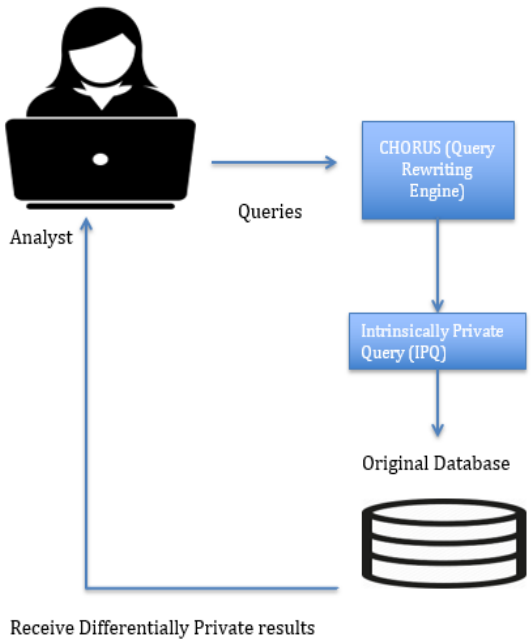
II. METHODS AND MATERIAL

IMPLEMENTATION OF CHORUS FRAMEWORK IN DP

CHORUS is a framework for data analysts in real world by ensuring privacy preserving of data. Chorus can be used by any non-expert and the data analyst doesn't need to understand differential privacy. It

automatically enforces DP for SQL queries. In addition to this, Chorus provide a modular design to support wide variety of mechanisms and it can be easily integrated with existing data environments.

Chorus take a different approach to enforce DP by Query Rewriting Technique without changing the original database. Chorus takes in the query the analyst's writes and produces a query called IPQ (Intrinsically Private Queries). This IPQ can be run on the original database to receive Differentially Private Results.



III. RESULTS AND DISCUSSION

PREDICTIVE DISEASE MODELLING USING MACHINE LEARNING ALGORITHM-RANDOM FOREST

The Differentially Private (DP) results of health data can be used to predict the disease risk of each individual. Supervised Machine Learning algorithms can be used to predict the risk of having a disease in near future. Out of all Machine Learning algorithms, Random Forest (RF) is having the highest accuracy.

RANDOM FOREST (RF) ALGORITHM

According to RF algorithm, a set of Decision Trees (DT) are grown and each tree votes for the most popular class of disease. Votes of different trees are integrated and a class is predicted for each sample data. This approach is designed to increase the accuracy of decision tree. This approach is an ensemble classifier composed of some Decision Trees and the final result is the mean of individual trees results

IV. CONCLUSION

Disease prediction has recently gained significant attention from data researchers across the globe without compromising patient's privacy. With the help of Chorus framework, we can filter only the differentially private data of individual. This DP data when fed into the Disease Risk Prediction Model of Supervised Machine Learning classifies and predicts Disease susceptibility of an individual without compromising data privacy.

V. REFERENCES

[1]. https://en.wikipedia.org/wiki/Fitbit
[2]. https://medium.com/georgian-impact-blog/a-brief-introduction-to-differential-privacy-eacf8722283b
[3]. https://arxiv.org/pdf/1809.07750.pdf
[4]. https://en.wikipedia.org/wiki/Machine_learning
[5]. https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm

Cite this article as :
Nitha V R, "Predictive Disease Modelling of Differentially Private Health Data", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 139-140, March-April 2020.
Journal URL : http://ijsrcseit.com/CSEIT206240