Factors Affecting Health in New York City

Anuj Bora
New York University
New York, USA
aab688@nyu.edu

Gowtham Goli
New York University
New York, USA
sgg308@nyu.edu

***Abstract*—Big Data is having an increasing role in healthcare with various applications. With our work, we are trying to establish correlation between water quality complaints, air quality complaints and garbage related complaints to the health related issues. This will help to identify whether water, air and sanitation are the major reasons for health issues and if they are, tackling those issues will lead to good health of the community.**

Keywords—analytics, big data, healthcare

I. INTRODUCTION

The analytic will extract health related requests from the NYC 311 service request dataset using MapReduce [1]. Then, the data will be partitioned according to boroughs and it will be compared with water quality, air quality and garbage related complains obtained from NYC's air quality dataset, water quality dataset and OATH ECB Hearings Case Status (which will be used to extract garbage complaints) to gauge the impact of those factors on health in boroughs of NYC.

The project will help to conclude whether the air quality, water quality and undisposed garbage is deteriorating health conditions in the examined boroughs.

We are expecting to observe a direct correlation between the water quality complaints, air quality complaints and garbage related complaints to the health related issues that are potentially caused due to food poisoning, indoor air quality and increased rodents in the respective neighborhoods. A direct proportionality, if could be established between these datasets would serve as a proof to the goodness of our analytic.

II. MOTIVATION

Big data analytics can play a major role and provide multiple benefits in health care. Electronic Health Records of patient treatments can be used to increase clinical knowledge and support clinical research. Apart from this, sensors produce lots of data. This data can be worked on to produce useful results. Also, internet-of-things can make data acquisition easy and thus making application of big data analytics in health care more streamlined.

So, in all, historical health data provides important insights in future health and that data can be used to improve clinical research which will foster health care resulting in better health service.

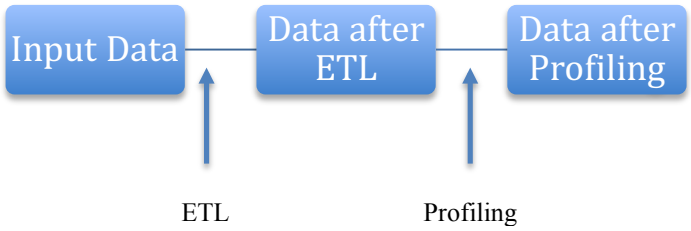
III. RELATED WORK

Big data is having increasing role in health [2] care and it is improving clinical research. The number of publication related to big data have been growing exponentially in recent years. Size is not the only factor that constitutes big data. Big data along with having large size also has heterogeneity and variety in the data. Apart from these, value, volume, velocity, veracity and variability are the other factors which define big data.

A practical and hands on methodology is proposed by Wullianallur Raghupathi, Viju Raghupathi [3]. In Step 1, the interdisciplinary big data analytics in healthcare team develops a 'concept statement'. In Step 2, questions like What problem is being addressed? Why is it important and interesting to the healthcare provider? etc. are undertaken. Step 3 includes platform/tool evaluation and selection such as AWS Hadoop, Cloudera, or IBM BigInsights etc. In Step 4, the models and their findings are tested and validated. Various real life successful applications of big data analytics in healthcare are then discussed.

IV. DESIGN

The pipeline of the operations which will be performed on the data can be represented in a diagram as below.



V. RESULTS

(Future… In this section, can describe: Your experiment setup/issues with data/performance/etc.)

VI. FUTURE WORK

The analytics can be further worked on by including more datasets related to other factors which may lead to health issues.

Also, more historical data can be collected and worked on which will gives us more complete idea on prioritizing the factors which affect health.

VII. Conclusion

(Future… One or two paragraphs about the value/accuracy/goodness of your analytic.)

Acknowledgment

(List any people, companies, organizations that you would like to thank.)

References

[1] J. Dean and S. Ghemawat. MapReduce: Simplified data processing on large clusters. In proceedings of 6th Symposium on Operating Systems Design and Implemenation, 2004.

[2] Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang

[3] Health Information Science and Systems, 2014, Volume 2, Number 1, Page 1 Wullianallur Raghupathi, Viju Raghupathi.