BIG DATA ANALYTICS FOR EFFICIENT WASTE MANAGEMENT

Parag Kedia[1]

[1]*Computer Engineering (B.E.), P.I.C.T, Pune, Maharashtra, India*

Abstract

This paper focuses on performing Data Analytics and learning the data trends using a few Machine Learning algorithms. Basically the data gathered from all dumpsters, their capacity of how much it is filled will be captured and we can perform Data Analytics and gain some insight from the collected data. Some of the supervised and unsupervised methods like KNN and Density Estimation can be applied on data and actions can be taken for future operations to be performed on the data. This will help in Value Creation in the IoT Industry which will bring benefits to the owner or to the society.

Keywords-*Big Data Analytics, Internet of things, Machine Learning, Cloud computing.*

----------***----------

1. INTRODUCTION

Since we have first appeared on this planet, humans have been producing waste. At the start of the 21st century most world's waste is still treated as it always has been, just thrown away more often and not, ending up as a landfill. We can't go on like this and it's just not sustainable. New approaches and technology can change the way we handle our waste. This is good news not only for us but for the whole planet. Many methods of municipal solid waste management have been developed and many of them have potentially significant adverse impacts. A Waste Management Sensor deployed on top of trash-bin can be used to get the capacity information of the trash-bin[1]. The data gathered can be sent to the Cloud resulting in enormous quantity of data from thousands of Sensors.

Big Data is about the processing and analyzing large data repositories with many available tools in tolerable amount of time. Data gathered is of enormous quantity containing various types of information about capacity, location, time/date of pickup, pickup vehicle assigned, etc. This can be helpful in predicting exactly when the trash-bins are getting filled and when to schedule pickups so as to avoid higher transportation and fuel cost, labour cost, etc.

2. BIG DATA AND THE INTERNET OF THINGS (IOT):

Big data and the Internet of Things (IoT) are leading in a new generation of quantitative analysis, leaving behind historical business intelligence approaches and solutions. Candidly, they are using the new data and new math to generate insights that increase precision, predict more accurately and enable greater personalization.

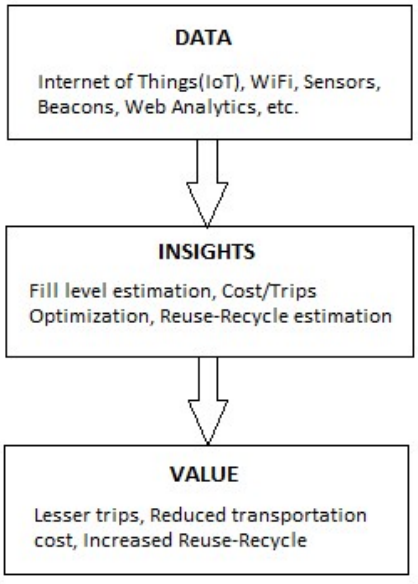
As a result, companies find that navigating through the data and creating insights from it is far easier than focusing on the right insights and applying them to unlock its full potential value. It helps in improving operational efficiency, effectiveness and competitiveness.

3. MOTIVATION

Big Data consists of high Volume, Velocity, and Variety(the **3 "Vs"**) of information demanding cost-effective and innovative data processing for gaining insights and decision-making. Here, a fourth "V" can be added to signify the data's Value. A primary challenge is discovering the various sources, quantity, and quality of waste resource requirements. To overcome this information barrier, we propose the big data approach. The value typically shows up as a result of a combination of IOT applications with new process introduction and accumulates over time. It is difficult to quantify value due to complex processes and multiple interactions. Fact is that value is the key element finally asked by the project stakeholders or owners.

Therefore, by applying data analytics we can analyze that as more waste is generated within the same area, it will require more waste pickup vehicles for dumping. So, the days of pickups can be minimized thereby saving fuel, labour cost and time associated with it.

Architecture Diagram



4. METHODOLOGY

As with most big data analysis first step in this process is to understand the current situation, analyze it, and then identify intervention measures. Various activities associated with it are identifying datasets, presenting data and aggregating it for decision making.

4.1 Dataset

It consists of many entries:
• Capacity of trash-bin
• Time and date of collection
• Collection address
• Assigned pickup vehicle and
• Final waste disposal site.

4.2 Data Presentation

The data can be presented in form of Graphs/Tables like:

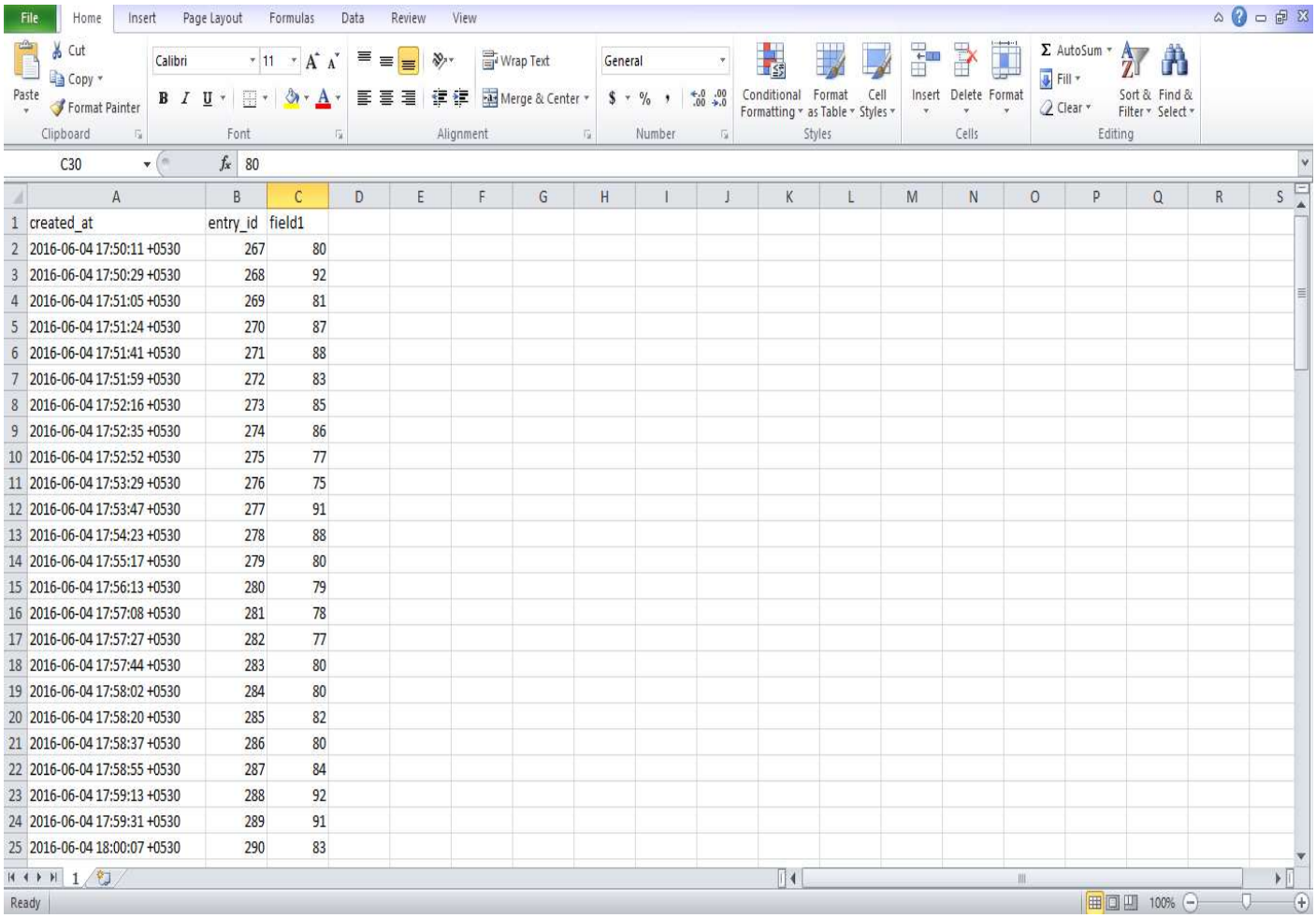
- Solid waste from different locations and proportion of waste components from each location (Tables/Pie charts)
- Time-series data and projections (Tables/line charts)

4.3 Aggregation of Waste Data

The data from various locations can be aggregated in order to identify when a higher number of active wastes are generated in particular week and also which part of the city produces what kind of waste.

5. IMPLEMENTATION FRAMEWORK

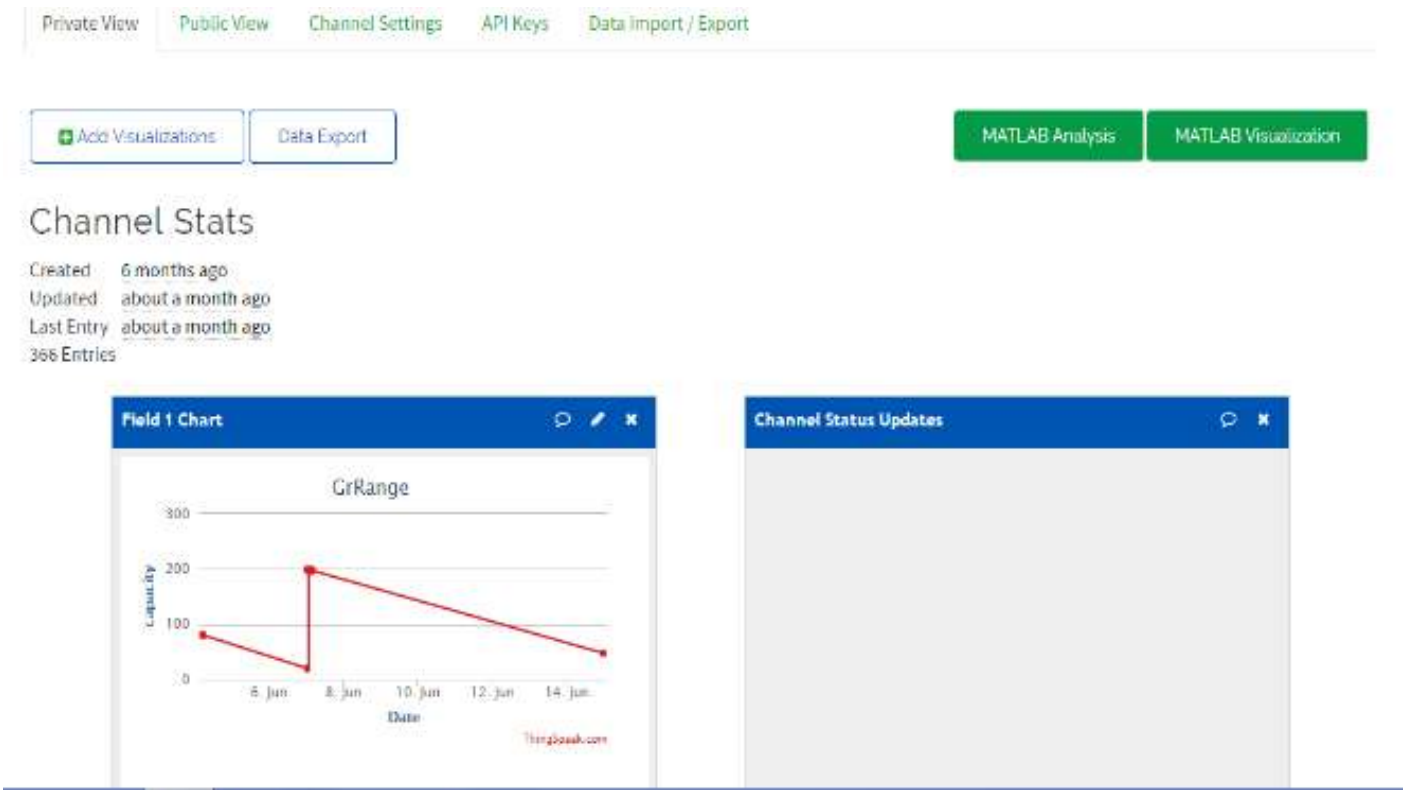
As shown in the diagrams below, the cloud consists of single channel along with its field. Many available options such as Data Import/Export option can be used to easily extract the data. The data can be exported in any of the JSON, CSV or XML format. Once the data is extracted, it can be used for applying analytics on it. The data will contain various attributes associated with it like- date/timestamp, tag associated with it and the feed values. It can be seen as-



	A	B	C
1	created_at	entry_id	field1
2	2016-06-04 17:50:11 +0530	267	80
3	2016-06-04 17:50:29 +0530	268	92
4	2016-06-04 17:51:05 +0530	269	81
5	2016-06-04 17:51:24 +0530	270	87
6	2016-06-04 17:51:41 +0530	271	88
7	2016-06-04 17:51:59 +0530	272	83
8	2016-06-04 17:52:16 +0530	273	85
9	2016-06-04 17:52:35 +0530	274	86
10	2016-06-04 17:52:52 +0530	275	77
11	2016-06-04 17:53:29 +0530	276	75
12	2016-06-04 17:53:47 +0530	277	91
13	2016-06-04 17:54:23 +0530	278	88
14	2016-06-04 17:55:17 +0530	279	80
15	2016-06-04 17:56:13 +0530	280	79
16	2016-06-04 17:57:08 +0530	281	78
17	2016-06-04 17:57:27 +0530	282	77
18	2016-06-04 17:57:44 +0530	283	80
19	2016-06-04 17:58:02 +0530	284	80
20	2016-06-04 17:58:20 +0530	285	82
21	2016-06-04 17:58:37 +0530	286	80
22	2016-06-04 17:58:55 +0530	287	84
23	2016-06-04 17:59:13 +0530	288	92
24	2016-06-04 17:59:31 +0530	289	91
25	2016-06-04 18:00:07 +0530	290	83

The data sent to the cloud can be seen on the channels, something as shown below-

This data can be used for further analysis.



6. SOLUTION FRAMEWORK

As the data gathered from the sensors is enormous, various business insights and correlations can be gained in order to create value from the Big Data. Various machine learning algorithms can be applied depending on availability of target variables. Some of them are given below:

6.1. Supervised Learning and its Implementation

In this method, target(dependent) variable is predicted from a given set of predictor(independent) variables. Using these set of variables, a function mapping inputs to outputs is generated. This process continues until the model achieves a desired level of accuracy on the training data. Examples of this method are: Linear Regression, KNN, Decision Tree, Naive Bayes, SVM, etc.

KNN (K-Nearest Neighbors) algorithm can be used to identify the days on which the particular Trash-bins are getting filled. Using this algorithm, the capacity data can give information about the days on which they are getting filled in every week. These days requiring pickups in every week can be estimated and vehicle can be assigned for pickup. As not all trash-bins will require daily pickups, this will help in minimizing the number of Pickups associated with the collection of garbage. Also, this will significantly help in reducing the cost of transportation by minimizing the Fuel cost and Labour cost. The time to invest in these activities will also get reduced as there will be less number of pickups for that area on that particular day.

Knowing the traffic condition can also help in scheduling pickups at an appropriate time. The traffic data can be analyzed in order to get best suitable time for the pickup. This will help in minimizing the problem caused by the garbage to the commuters travelling on road when the pickup vehicle carries garbage to the dumping ground. Also, less traffic will minimize the cost of fuel.

6.2. Unsupervised Learning and its Implementation

In this method, none of the target or outcome variables are present for prediction. The data is clustered into different classes. The categorization is done so that the algorithm differentiates correctly between different variables in the data. Examples of this method are: K-means, Apriori algorithm, Density estimation, etc.

Various locations such as Industrial, residential area, healthcare area, construction site, etc. are present in the city. The waste can be of many types: metal, plastic, wood, paper, sand, syringes, injections, household waste, paint, sand, debris, etc. These items can be reused or recycled. Density Estimation will try to understand the areas where specific kind of waste is found in large quantities. As we do not know which types of waste is collected from a particular area, we can identify the types of waste and the quantity from where they are generated the most. Thus in order to meet specific needs, by knowing the waste type and its quantity we can collect it easily from that particular location for reusing and recycling.

7. CONCLUSION

Thus by identifying some strong points in data use, we can plan waste management strategy and more effective ways of data collection and analysis. It will help to evaluate options in order to make informed decisions (using some supervised and unsupervised learning methods). It discusses a few methods for data analysis that help in reducing the transportation cost i.e. fuel and labour cost along with the time spent in this process. The sources where specific type of waste can be found in large quantities will be identified and this will help in reuse and recycling of that waste in order to meet specific targets.

REFERENCES

[1]. Efficient Garbage Collection Using WSN Chaitanya More, DarshanMestry, Parag Kedia, Reshma More.
[2]. A big data analytics approach to develop industrial symbioses in large cities Song Bina*,Yeo Zhiquana, Low Sze ChoongJonathana, Derek KohJieweia, Denis Kurleb, Felipe Cerdasb, Christoph Herrmannb.
[3]. Municipal Solid Waste Management: Data Analysis and Management Options VaclovasTricys, SkaidrėZickiene, Ala Kovieriene.
[4]. Effective Use of Data in Waste Strategy Planning In England Christine I. Thomas.
[5]. Comparative Analysis of Solid Waste Management in 20 Cities David C. Wilson, LjiljanaRodic, Anne Scheinberg, Costas Velis and Graham Alabaster.
[6]. Analysis and overview of industrial solid waste management in Kuwait Jasem M. Alhumoud, Fatima A. Al-Kandari.
[7]. Internet of Things- Converging Technologies for Smart Environments and Integrated Ecosystems OvidiuVermesan, Peter Friess.

BIOGRAPHIES

Parag Kedia, Computer Engineering (B.E.), P.I.C.T Pune, Maharashtra, India.