

Necessity of Big Data and Analytics for Good e-governance

Bhavya Patel¹, Sudipta Roy², Debnath Bhattacharyya³ and Tai-Hoon Kim⁴

^{1,2}*Department of Computer Science and Engineering, U.V.Patel College of Engineering, Ganpat University, Kherva, Mehsana-3840112, Gujarat, India*

³*Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Visakhapatnam 530049, AP, India*

⁴*Department of Convergence Security, Sungshin Women's University, 249-1, Dongseon-dong 3-ga, Seoul, 136-742, Korea*

bhavya.bda1407@ict.gnu.ac.in, sudiptaroy01@yahoo.com, debnathb@gmail.com, taihoonn@daum.net

Abstract

Big Data is the most trending topic in the computer science field. This big data analysis is booming field in present days and used to get the insight in the particular domain. Big data analytics is used to solve many problems and get the predations of many upcoming problems. Big Data refers to the huge dataset which is not managed and handle by old traditional methods. Big Data plays a very important role in effective management of e-governance. In this paper we collect the kishan call center (KCC) data as the data is in semi structured format i.e in JavaScript Object Notation (JSON) format so we used mongoDB to store. For analysis purpose before using the data we had converted it into the structured format. After converting into the structured format we created charts for better understanding. This paper focuses on how the big data and analysis is helpful for good e-governance. This big data is useful to get the insight and predictive analysis that gives good e-governance.

Keywords: *Big Data, Data Analytics, E-Governance, Public Sector, Structured Data, Unstructured Data*

1. Introduction

Congratulations! New technologies contribute a lot in increasing the data day by day. Social media and the audio and video data and other unstructured data contributed a lot in the data growth. When generated data crosses the limit of GBs it falls into the big data category. As per definition of Big Data, data must at least follow the 3 V's Volume, Velocity, and Variety. Approximately of 2.5 quintillions of data are generated by users per day [1] and is huge data to manage from the management point of view. So Big Data is the technology to manage, store and handle this huge amount of data having various variety generating very rapidly daily. Analyzing this huge data is also challenging and can be solved by various tools like Hadoop, MapReduce, and Pig. Importance of e-governance is very important for the developing countries like India which lead to the creation of large data set within country. Indian government is also planning to implement the concept of smart cities which in future will contribute large chunk of data. Analyzing these huge data in turns give the result which is helpful for prediction of future projects. The source to collect the data are social media, device or sensor generated data, internal and external data, surveys done by government and other private company and many other methods. For e.g.: from social media like twitter, survey data done by government and private companies like during census counting or data collected by private company for exit poll. Following Figure 1 shows the basic overview of data sources analysis of big data.

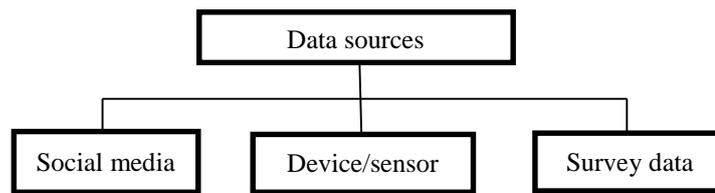


Figure 1. Different Data Sources

Any government entity or organization that uses the information and communication technologies to operate their business and services is known as e-government [2]. E-governance is the collection of services of the government provided to the people [2]. In India national e-governance plan NeGP is implemented by Department of Electronics and Information Technology (DeitY) and Department of Administrative Reforms and Public Grievances (DARPG)[2]. The vision of NeGP is “Make all Public Services accessible to the common man in his locality, through common service delivery outlets and ensure efficiency, transparency and reliability of such services at affordable costs to realize the basic needs of the common man” [3]. For better governance it is very important know the trend, pattern and the sentiment of the public for the policies or schemes or rules they have formed. Generally, the government follows the objectives for e-governance [4]: i) To build an informed security, ii) Increase government and citizen interaction, iii) Encourage citizen participation, iv) Bring transparency in governing process, v) Make government accountable, vi) Reduce cost of governance and vii) Reduce the reaction time of government.

There are typical four types of model for e-governance and they are: i) G2G(Government to Government), ii) G2C(Government to Citizens), iii) G2B(Government to Business), and iv) G2E(Government to Employee).

In this paper we have discussed the benefits of the big data and analytics in the e-governance. Using the analytics generates the reports and this report is analyzed to take action on the new reforms or scheme or plans. There are methods like data mining to get the insight from the data which is helpful for betterment of e-governance.

2. Big Data Analytics for E-Governance

Government of India (GoI) uses the information technology to deliver the information in the field of health care in rural areas. Also government is delivering the information related to the government scheme to the public of rural area. Indian government has started this project with an aim and objective of removing of duplicate entries from the government records and to provide high quality low error database, generates the reports automatically or regular basis. Use of data electronically for the future purpose and also some data is electronically available to the people. The traditional e-governance model is based on the customer satisfaction, process the information and e-governance as tool for government. Figure2 is the e-governance working framework using traditional model [5].

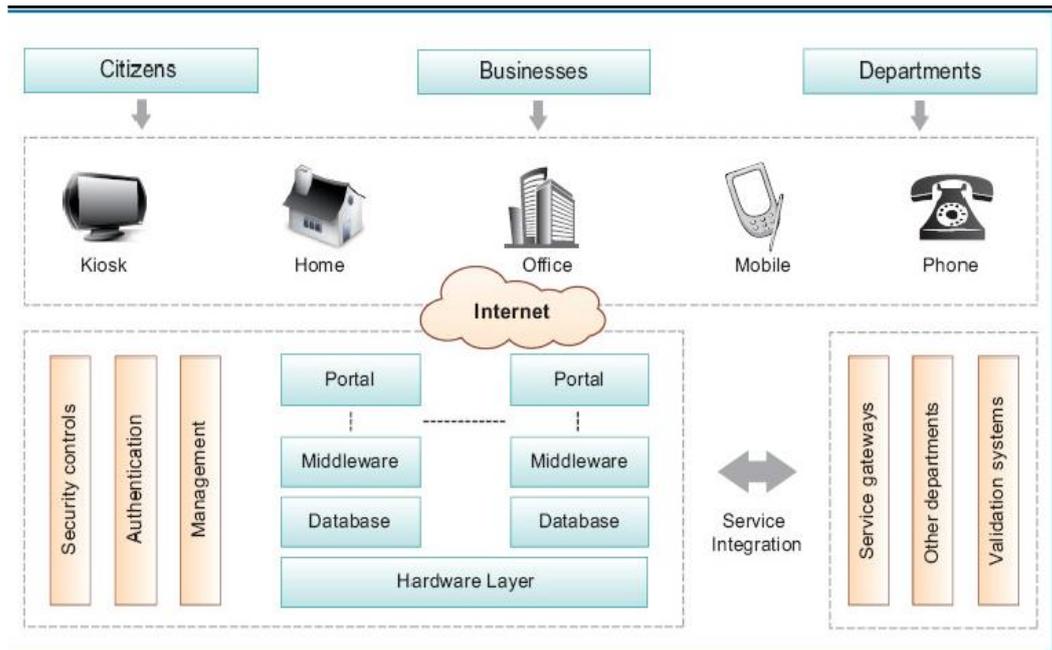


Figure 2. e-governance Framework using Traditional Model [5]

Matthias Finger in his paper[6] suggest the new model for e-governance and compare the traditional method customer satisfaction , process the information and e-governance as tool for government with the newly proposed method. Following Figure 4 is the suggested new model by National e-Governance Plan (NeGP) in network for information and computer technology (NICT).

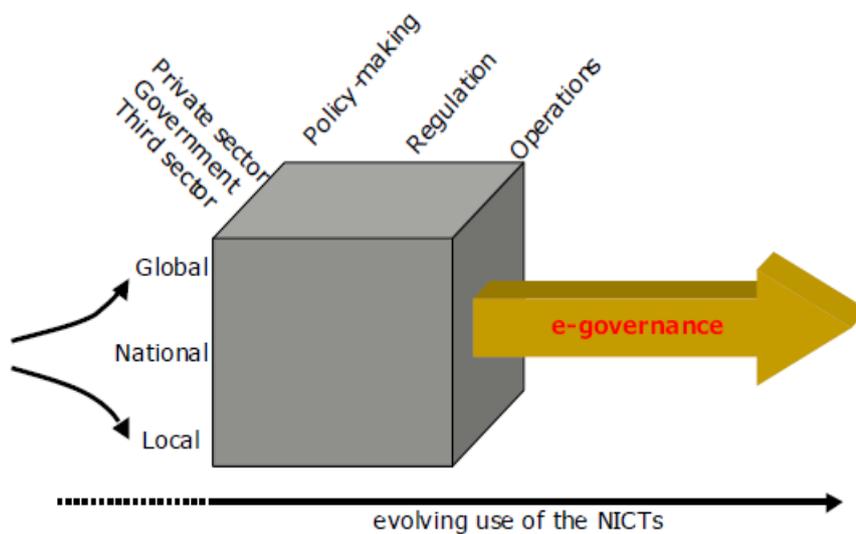


Figure 3. NICT Model in e-governance

As per this NICT model there is need of policy making, regulation and operations. Before taking the huge decision one should analyzed the previous data and upon the result one should take the decision. To get better understanding of behavior of data one should use big data analytics. If the decision making is done big data analytics will surely result into better decision. Indian government uses big data and analytics for implementing various policies

and to make change and reforms in existing system. The biggest contributor of big data in India is the 'Aam ADHmi ka AdhikAR' as 'AADHAR' scheme Unique Identification Authority of India (UIDAI) which give the unique. 100 crore Unique ID number has been generated as on 4th April 2016 [7]. This AADHAR number are used and linked with bank account, Liquefied Petroleum Gas (LPG) connection, 'Mahatma Gandhi National Rural Employment Gurantee Act (MANREGA)' job cards, 'Indian railway catering and tourism corporation limited and passport (IRCTC)', and passport. Now this data is huge and is link with many other schemes to maintain we need big data and to get the insight we need the big data and analytics technology. Similarly, the digital India initiative of an India government aims at ensuring the government services are made available to citizens electronically by reducing paper work which in turns make the process fast. There are many other sources from which data is generated in huge amount and which follows 3'Vs of Big data. Following diagram shown in Figure 4 suggest the main sources of data in India.

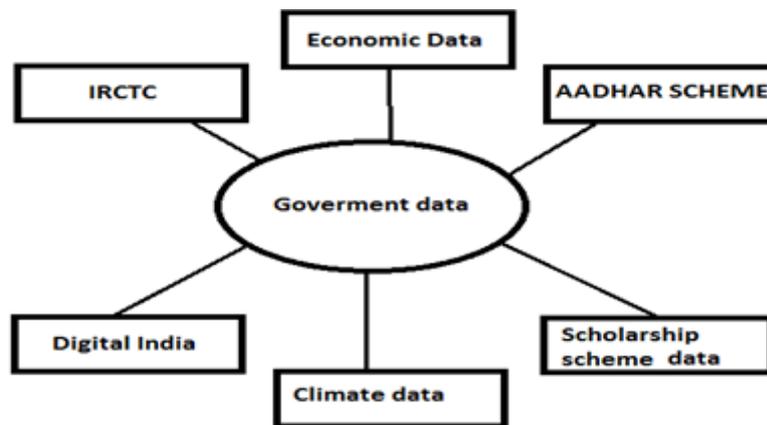


Figure 4. Main Sources of Data in India

We get the benefits from analytics like predictive policing, evolving new market, increase operational frequency, and better consumer services.

3. Storing of e-governance Data

Government and its organizations are the biggest producers of data in any country. This is the big question that how to store the different type of data. In e-governance there are many different types of format of data. It is difficult to store it in the RDBMS as it need predefined database schema. As the data structure of government data is always changing. To solve this problem one should create different database structure of every format of the data coming from the public. This is very hard to handle and maintain the database, also it proves very expensive. When the data increase in huge amount to petabytes it is very difficult to retrieve the data and also it takes very long time to give the result. So it is not feasible to store the data in the RDBMS. But the transaction data must be stored in RDBMS as it provides ACID properties. Another way to solve the storage problem of data is to store the data in Not Only Structure Query Language (NOSQL) Database. In NOSQL there are main 4 types of data model, they are (i) document-oriented, (ii) columnar database, (iii) key-value database, and (iv) graph database. The reason to select the NOSQL database is that NOSQL provides flexible schemas.

Document-oriented database is best for storing the semi structured data like JavaScript Object Notation (JSON) format file. It uses the whole document structure to store the data. It doesn't need any predefined schema and one can add new attributes when required. We have given an example of format that can be store in NOSQL databases is shown in Figure 5 below.

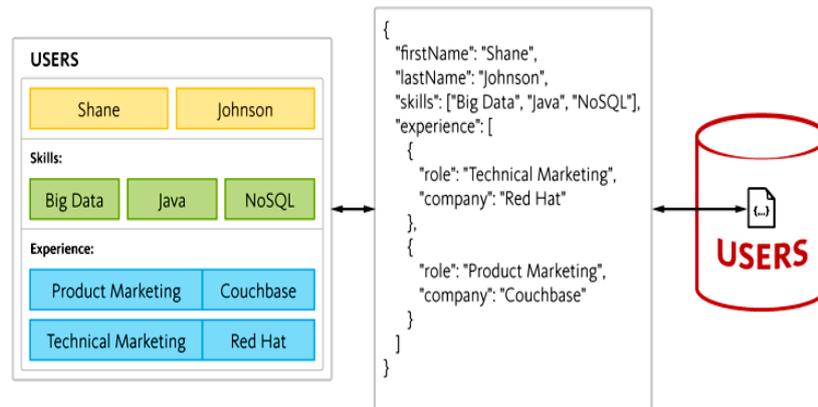


Figure 5. Document Database (couchbase)

NOSQL documents database have no need of joins as it stores the data in single document and is shown in Figure 5. As most the data in government are in semi structured data so document-oriented database is the best suited for the e-governance. For *e.g.*: census data will be stored and maintained well in document oriented model of NOSQL. Also the health record of public in hospital is also well maintained in NoSQL document oriented model. One example of document database has been shown in Figure 6 below within box.

```
{
  "_id": "abc_123",
  "aadhar": "1234 45678 9090 9090",
  "data": {
    "Health": {
      [
        {
          "Date": "10/08/2011",
          "Physically_handicapped": "NO"
          "Any_Allergy": "Tonsils"
        },
        "Medicine": {Date: "09/28/2011",
          Amount: "12"
          Quantity: "5"
          frequency_per_day:"1"
        }
      ]
    }
  },
  "formdef": "fs_forms_medication_1",
  "metadata": {
    "changed-by": "abc 123",
  }
}
```

The other one which is also very helpful is graph database. It is used to find the relations between the entities and use those relations for getting more insight in the data. For example, if father or mother is having some allergy than there may be the chances of having the same allergy to his child. This we can plot the using the graph database by relating child and father/mother entity by parent child relation. We can take early precaution in this case and may help to cure that allergy and it is shown in Figure 7.

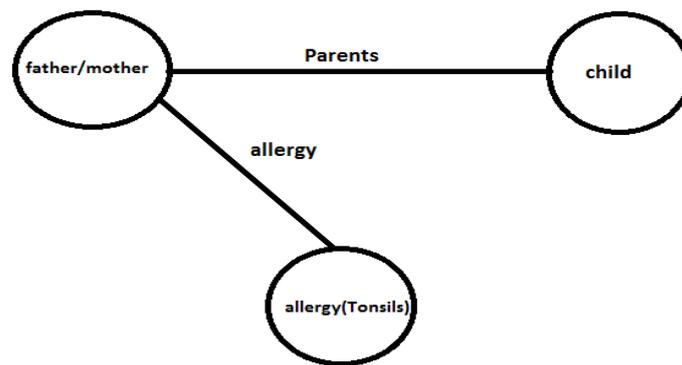


Figure 7. Graph Database showing Parents Child Relation along with the Allergy Parents have

4. Different Methods for Big Data Analytics

This section of paper discussed about the different approaches and methods for big data analytics. There several methods out of that most commonly highly used are: i) data mining, ii) text analytics, and iii) machine learning. Data mining is the technique or the process of analyzing and managing the data from the different perspective and to get the useful insight by summarizing. In short data mining is the process to correlates the fields in large datasets. Text analysis is a technique to analyzing the specific text or finding some pattern from huge data sets of file. Text analysis can be done on social network data, web data, and documents. Machine Learning-is a technique which guide the machine to learn to do analysis by using various algorithm. It focuses on the dynamic programs that change as per the new data. It helps in performing the analysis by learning the data analysis algorithm. In machine learning it learns more and more and become more accurate by experiences. The machine learning algorithm applied in the mind reading wizard [8, 9, 10, 11, and 12] which neither take any input nor read. It will communicate by the users thought. It is one of good example of machine learning. There are many more methods like sentiment analysis, regression analysis, classification tree analysis, Association rule learning, genetic algorithms, and Social network analysis.

5. Result and Discussion

Here one use case related to e-governance has been shown in Figure below. This use case is related to the farmers asking query to the kishan call center (KCC). This data is taken from the website 'data.gov.in'. The data is in semi structure format that is in JSON format. So we stored this data in NOSQL database using MongoDB. Here is the screenshot of our MongoDB data has been shown in Figure 8 below.

```
{
  "Season": "RABI",
  "Sector": "AGRICULTURE",
  "Category": "Others",
  "Crop": "Others",
  "QueryType": "Weather",
  "QueryText": "WEATHER REPORT",
  "KCCAns": "THERE IS NO RAIN WITHIN 24HOURS AS PER IMD REPORT",
  "StateName": "ODISHA",
  "DistrictName": "SAMBALPUR",
  "BlockName": "RAIRAKHOL",
  "SubmitDate": "1/31/2016 4:58:50 PM"
},
{
  "Season": "RABI",
  "Sector": "HORTICULTURE",
  "Category": "Vegetables",
  "Crop": "Tomato",
  "QueryType": "\tPlant Protection\t",
  "QueryText": "LEAF SPOT",
  "KCCAns": "SPRAY SAAF 2GMS/1LITRE OF WATER",
  "StateName": "ODISHA",
  "DistrictName": "SAMBALPUR",
  "BlockName": "KUCHINDA",
  "SubmitDate": "1/31/2016 5:08:14 PM"
},
}
```

Figure 8. KCC Data in JSON Format Stored in MongoDB

This data contains only information queries from Puri and Shambalpur districts of Odisha, India. Now we had done the analysis of this data and get the insight. We get the insight that from which region the numbers of queries are more, also in which season they the more queries. Using the insight get from this data or from the analysis the government can plan which type of campaign they have to start. Also we analyze for which crop the queries are more and from this analysis government start program which create awareness among the farmers related to the method of cultivation of that crop and what other measures should be taken. Here are some screenshot that we get from the analysis of data.

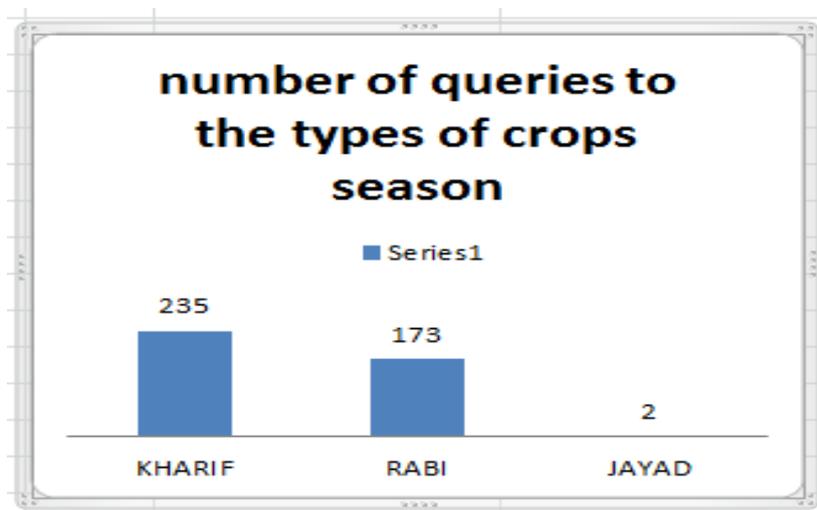


Figure 9. Number of Queries Related to the Season from KCC Data

Kharif crop are basically a summer crops which are cultivated during rainy season. From above Figure 9 we get to know that the farmers are having problems or queries or lack of knowledge related to kharif crop so government can setup their camp especially for kharif crops. Rabi crops are crops sown in winter and harvested in spring. Zaid or Jayad crops are crops which are irrigated and harvested in between of kharif and rabi crop season.

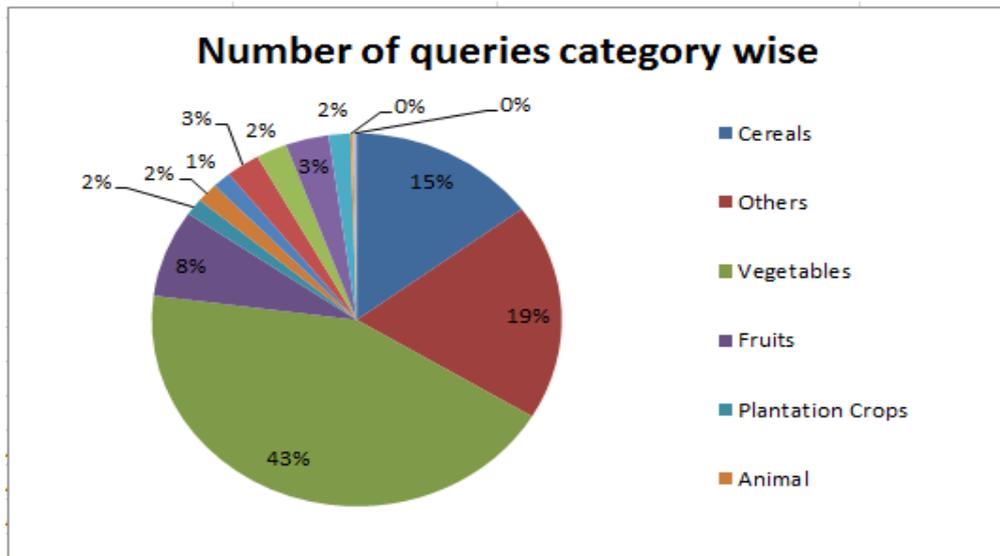


Figure 7. Analysis of KCC Data

From the Figure 10 we acquire the knowledge that the farmers from Puri and Shabmhalpur district are more related to vegetables farming. Also we get to know that the farmers are having the less knowledge or more queries related to vegetables crops. So government can plan some public programs to teach new methods for the good cultivation of vegetables crops. Now we know that the farmers from Puri and Shambhalpur district are facing the problems in vegetables crop cultivation, we can dig deeper in the data and get more precise insight by getting the knowledge that in which particular vegetable crop the farmers are facing problems.

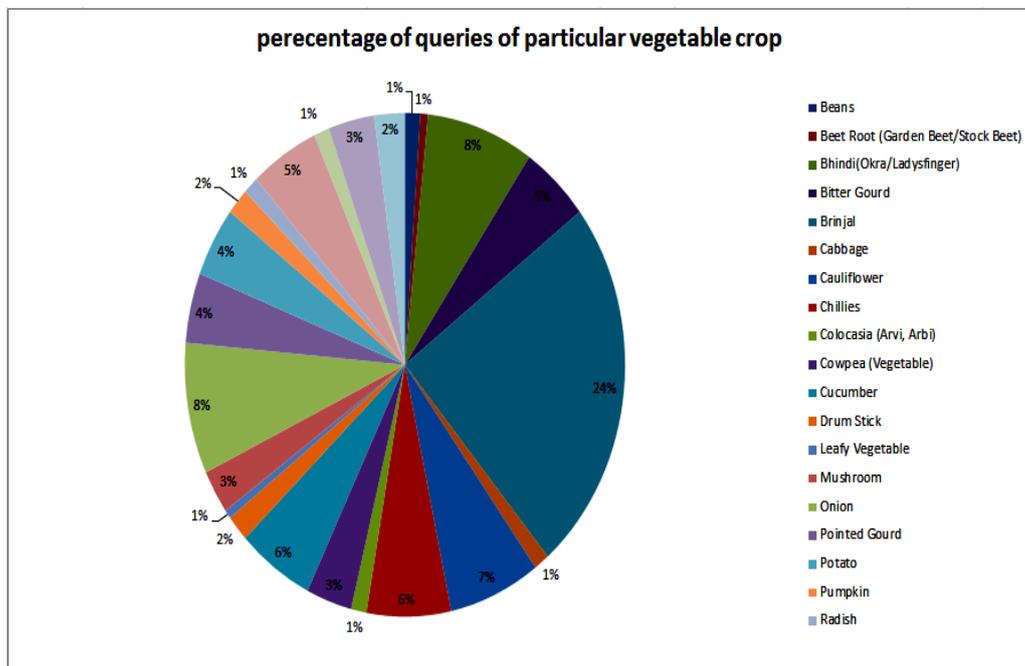


Figure 7. Percentage of Queries of Particular Vegetables Crop

Figure 11 shows a dig deeper in the data and to get the insight that in the vegetable crop the highest of the queries are related to the “*Brinjal*” crop and second place is taken by “*Bhindi (Lady’s Finger) and Onion*”. Using this insight government can plan to setup the

training camp for the cultivation of vegetables crops having more focus on the Brijanl, Onion and Bhindi(Lady's Finger). From this government came to know that in particular season in particular district which type of crop cultivation is in higher trend among the farmers.

6. Conclusion

If big data can be used in right way with some good analytics approach for e-governance domain it will surely improve the methods of work to be done. It will give rise to the new ideas and speed up the procedure. New reforms or scheme can be analyzed before implementing which make more chances of successful implementation of the scheme. Big data analytics in e-governance can give rise for effective services to the citizens. Implementing big data and analytics in e-governance can provide more effective idea. Although there are some challenges but research work are going in this field. The main advantage to big data and analytics in e-governance is to implement the schemes or services which really the public wants. Also government can efficiently setup their camps for different purpose we in turn beneficial to the public and create the awareness in public related to schemes or programs

References

- [1] C. Dobre and F. Xhafa, "Intelligent services for big data Science", Future generation computer systems, Elsevier, Volume 37, (2014), pp 267-281.
- [2] Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, Samir Belfkih, "An Overview of Big Data Opportunities, Applications and Tools," Intelligent Systems and Computer Vision (ISCV), IEEE, pp. 1-6, March (2015).
- [3] http://research.omicsgroup.org/index.php/National_e-Governance_Plan#Vision (accessed on:- 2/4/2017)
- [4] <http://indiaegovernance.blogspot.in/2008/03/objectives-of-e-governance.html> #Vision(accessed on :- 2/4/2017)
- [5] Mahesh Kumar Sharma, Kunwar Singh Vaisla "Towards Cloud Supported E-Governance Services Delivery Model," 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT), IEEE, pp. 537-539, (2014).
- [6] Matthias Finger, GaellePecoud, "From e-Government to e-Governance? Towards a Model of e-Governance," 3rd European Conference on e-Government Switzerland, pp. 52-60, (2016).
- [7] <http://pib.nic.in/newsite/PrintRelease.aspx?relid=138555#Vision>(accessed on :- 3/4/2017)
- [8] <http://www.csml.ucl.ac.uk/projects/pascal-outreach-drupal-6.20/?q=node/29#Vision>(accessed on :- 3/4/2017)
- [9] JagdevBhagal, Imran Choksi, "Handling Big Data using NoSQL," 2015 29th International Conference on Advanced Information Networking and Applications Workshops, IEEE, pp. 393-398, (2015).
- [10] Rhoda C. Joseph, Norman A. Johnson, "Big Data and Transformational Government," IT Professional IEEE, Volume 15, Issue 6, pp.1520-9202, (2013).
- [11] A. Cardenas, P. Manadhata, S. Rajan, "Big data analytics for security", IEEE Security & Privacy, vol. 11, no. 6, pp. 74-76, Nov (2013).
- [12] Rajagopalan M.R, Solaimurugan Vellaipandiyar, "Big Data Framework for National e-Governance Plan", 2013 Eleventh International Conference on ICT and Knowledge Engineering, IEEE, pp.1-5, Nov (2013).

