

E-ISSN: 2707-6644
 P-ISSN: 2707-6636
 IJCPDM 2020; 1(1): 35-40
 Received: 21-11-2019
 Accepted: 22-12-2019

Thanneeru Madhu
 M.Sc, Sri Venkateswara
 University, Tirupati, Andhra
 Pradesh, India

Analysis of android applications using big data analytics

Thanneeru Madhu

DOI: <https://doi.org/10.33545/27076636.2020.v1.i1.a.8>

Abstract

The conventional human needs were food, shelter and clothing. But as the digital world is emerging, these needs have been changing to smart phones, mobile data and applications. Digitalization has become one of the most important part of human life today. And to make sure it connected through the virtual world; applications play an important role. The development of android phones along with android applications has eased our lives. In the past decade there were lakhs of android applications developed creating huge amounts of data. The data presently generated by these applications is in terabytes and it may soon reach zettabytes and even yottabytes. Out of such huge data that gives the users enormous choices, the users are ambiguous in choosing which app actually meets their needs exactly. The solution to this problem is my paper where it classifies the applications based on the ratings, category and also use dynamic search. i can perform some more realistic handling of data operation through latest technology tool called Hadoop, a very famous Big Data analytical tool. Here I am using Map Reduce as a processing component and involving fault tolerant file system, HDFS (Hadoop Distributed File System) as storage component.

Keywords: Digitalization, virtual, android, ambiguous, analytical

1. Introduction

There are practically in excess of four billion cell phone customers at present. There's no distortion in saying that versatile application headway is the perfect spot to put your assets. As of now, most organizations are going for versatile applications for reaching a broad social occasion of individuals. The organizations are hoping to serve their necessities. Right now, are anticipating a better yield on beginning capital speculation. Regardless, it isn't as straightforward as it shows up. You will discover numerous portable applications in Application stores. Thus, the prerequisite emerges to get all reports with respect to the rapidly changing business segment designs. One must know about elements like customer lead to remain in the spotlight. You evaded one possibility, and your application will miss out lost in the swarm of a monstrous number of applications. Without a doubt, you dislike such a situation the slightest bit. In any case, in the event that you need to remain comparable to your rivals, you should refresh yourself constantly. You need to gather and supervise the flood of data which is a huge overpowering strategy truly. It is the spot Big Data accept a surprising activity in your market framework. Large Data is a full term for the advances and instruments used for a social event, sifting through, and separating tremendous and different educational lists. The utility of huge information advertising channels is to discover feature designs, customer tendencies, and covered plans. You may likewise think that its valuable is building up associations between different bits of information. The utilization of Big Data is impressively more in helping the association by making convincing promoting methods. There are a couple of essential data instruments in the market. Their usage is for choosing cost-adequacy, directing time, and examining data. It assists with bringing out valuable business bits of information.

2. Proposed System

The solution to this problem is our project where classify the applications based on the ratings, category and also use dynamic search can perform some more realistic handling of data operation through latest technology tool called Hadoop, a very famous Big Data analytical tool. Here I am using Map Reduce as a processing component and involving fault tolerant file system, HDFS (Hadoop Distributed File System) as storage component.

Corresponding Author:
Thanneeru Madhu
 M.Sc, Sri Venkateswara
 University, Tirupati, Andhra
 Pradesh, India

2.1 Algorithm

The random forest algorithm is an ensemble classifier algorithm based on the decision tree model. It generates k different training data subsets from an original dataset using a bootstrap sampling approach, and then, k decision trees are built by training these subsets. A random forest is finally constructed from these decision trees. Each sample of the testing dataset is predicted by all decision trees, and the final

classification result is returned depending on the votes of these trees.

The original training dataset is formalized as $S = f(x_i; y_j); i = 1; 2; \dots; N; j = 1; 2; \dots; M$, where x is a sample and y is a feature variable of S. Namely, the original training dataset contains N samples, and there are M feature variables in each sample. The main process of the construction of the RF algorithm is presented in Fig. 1.

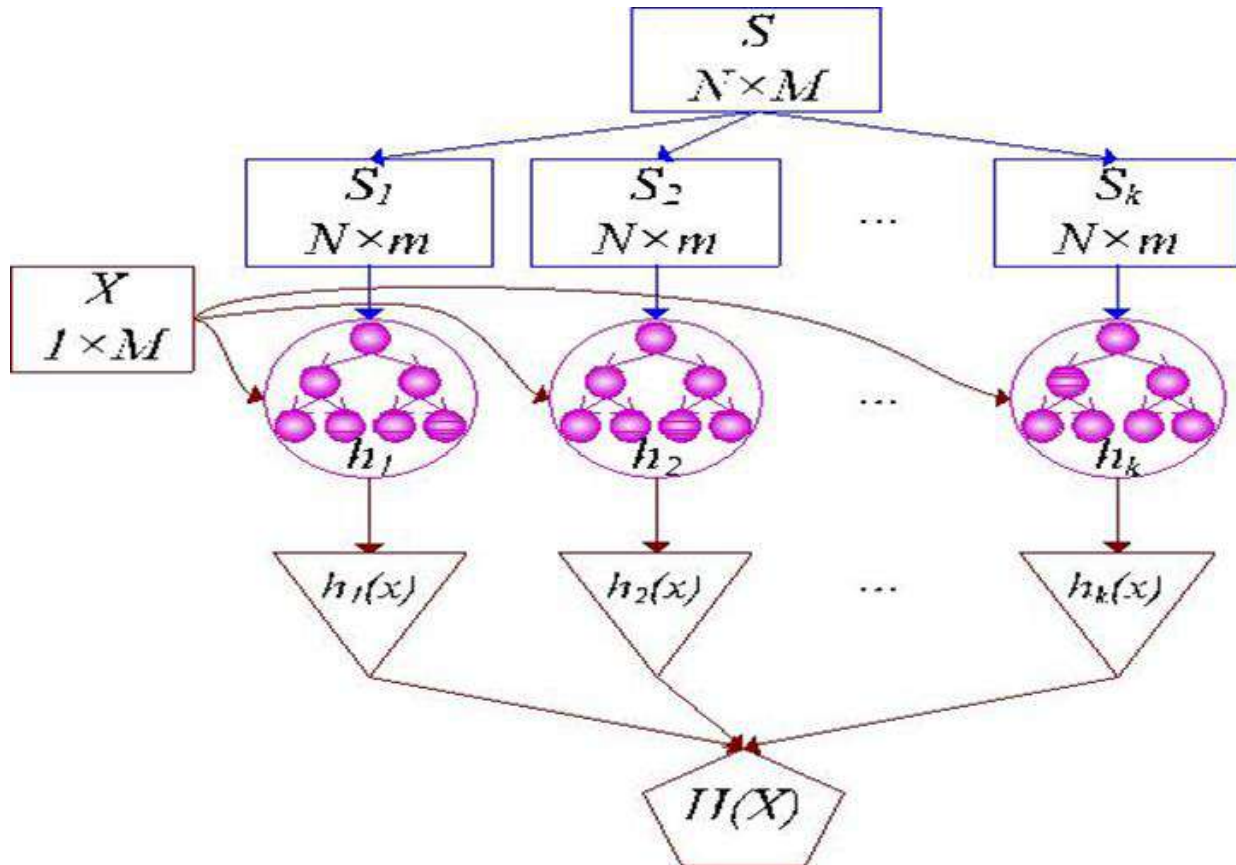


Fig 1: Process of the construction of the RF algorithm

$$S_{OOB} = f(OOB_1; OOB_2; \dots; OOB_k), S$$

Where $k \leq N$, $S_i \cap OOB_i = \emptyset$ and $\cup_{i=1}^k OOB_i = S$. To obtain the classification accuracy of each tree model, these OOB sets are used as testing sets after the training process.

Step 2: Constructing each decision tree model.

In an RF model, each meta decision tree is created by a C4.5 or CART algorithm from each training subset S_i . In the growth process of each tree, m feature variables of dataset S_i are randomly selected from M variables. In each tree node's splitting process, the gain ratio of each feature variable is calculated, and the best one is chosen as the splitting node. This splitting process is repeated until a leaf node is generated. Finally, k decision trees are trained from k training subsets in the same way

Step 3. Collecting k trees into an RF model.

The k trained trees are collected into an RF model, which is defined in

Eq. (1):

$$H(X; j) = \sum_{i=1}^k h_i(x; j); (j = 1; 2; \dots; m); \quad (1)$$

Where $h_i(x; j)$ is a meta decision tree classifier, X are the input feature vectors of the training dataset, and j is an independent and identically distributed random vector that determines the growth process of the tree.

3. Results and Discussions

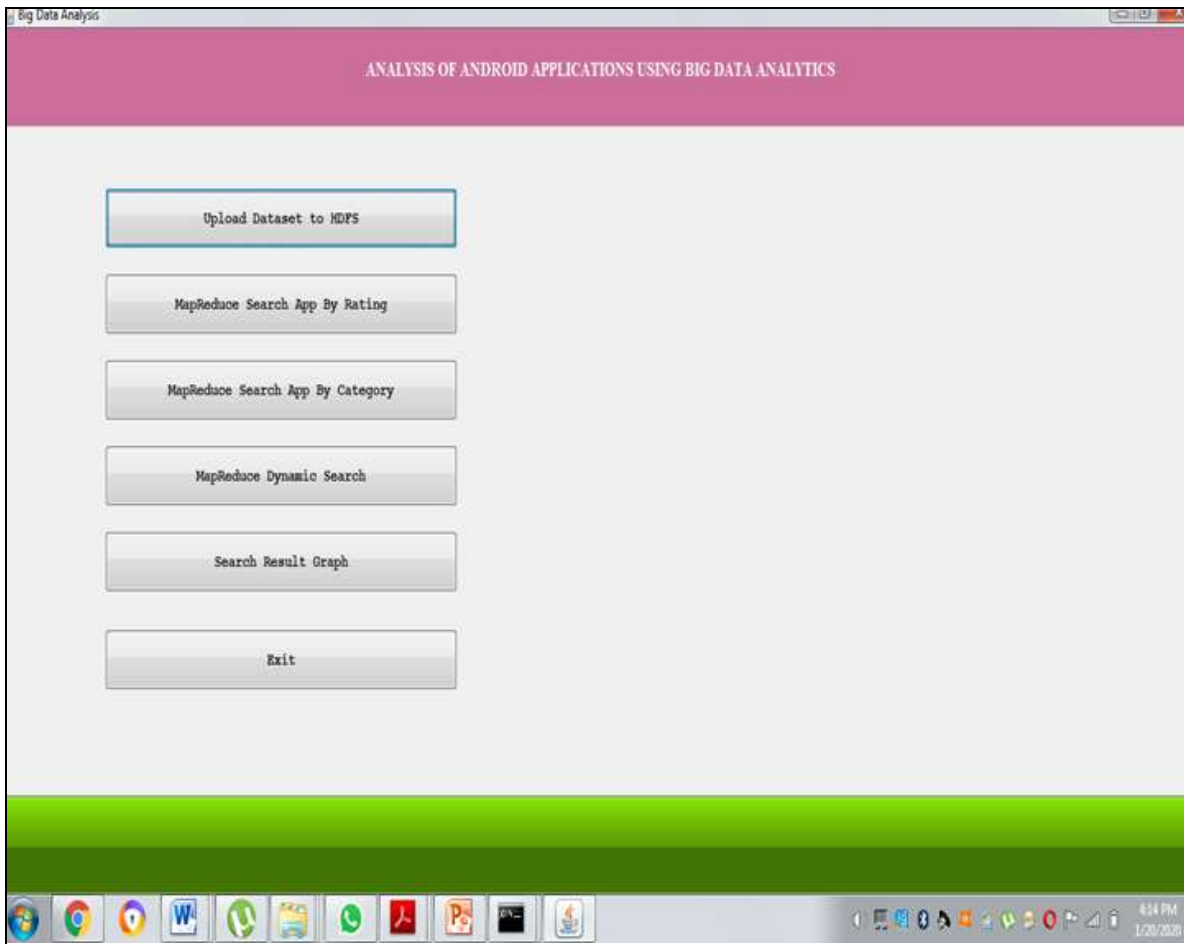


Fig 1: Home Page

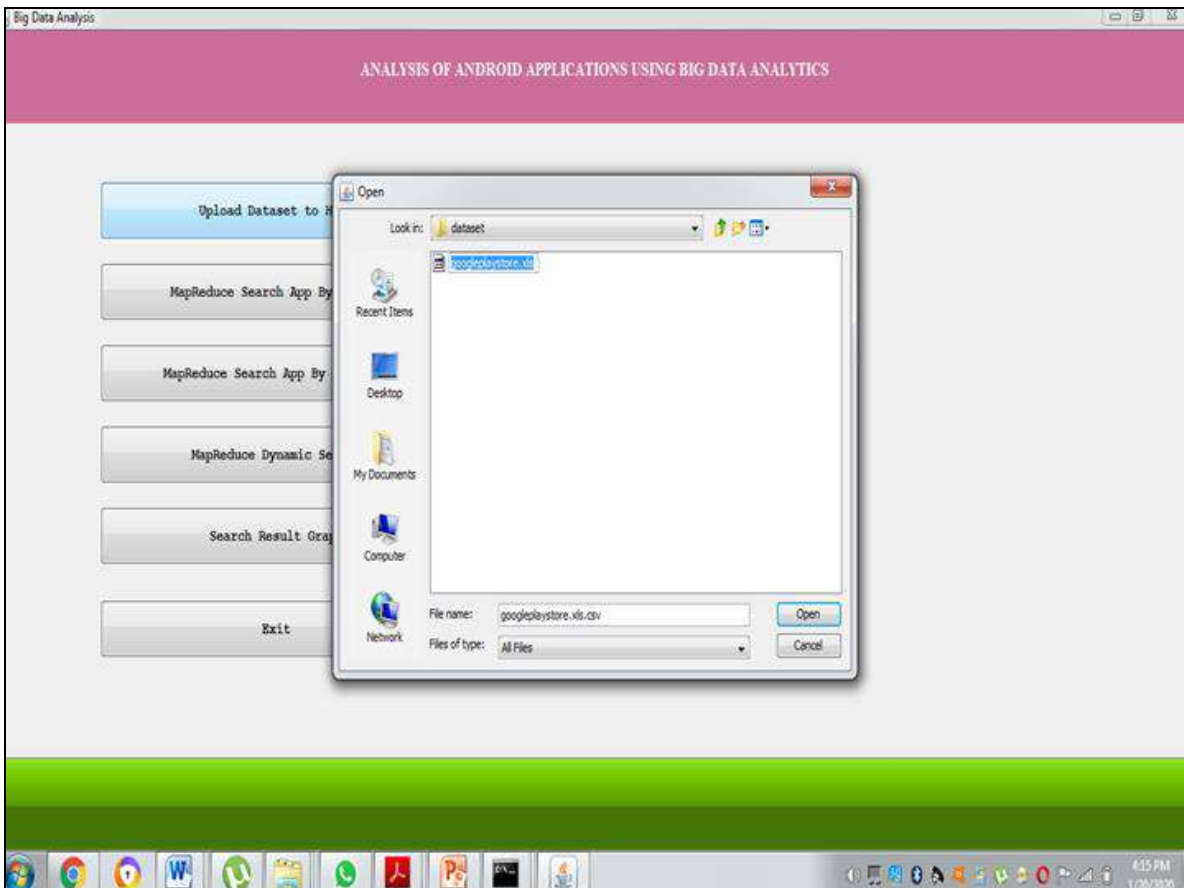


Fig 2: Uploading Dataset

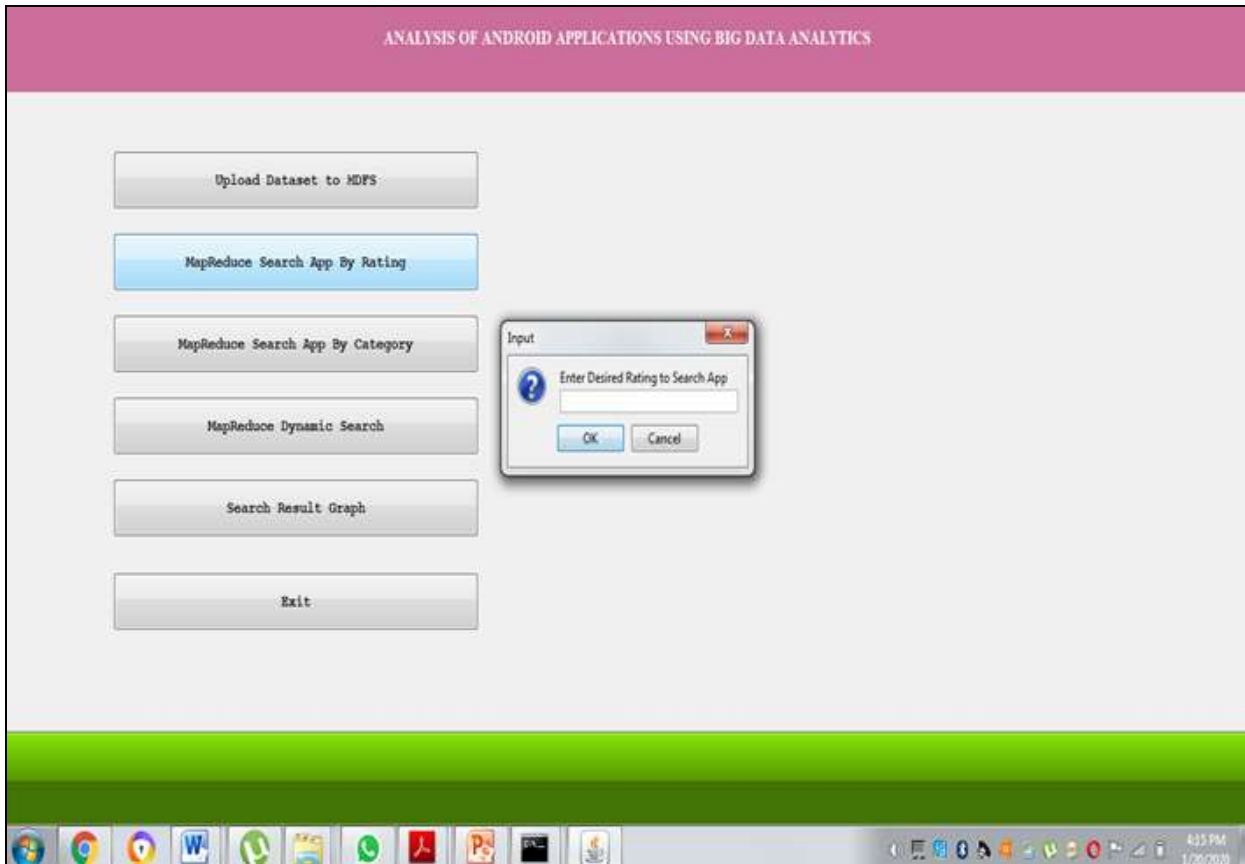


Fig 3: Search by Rating

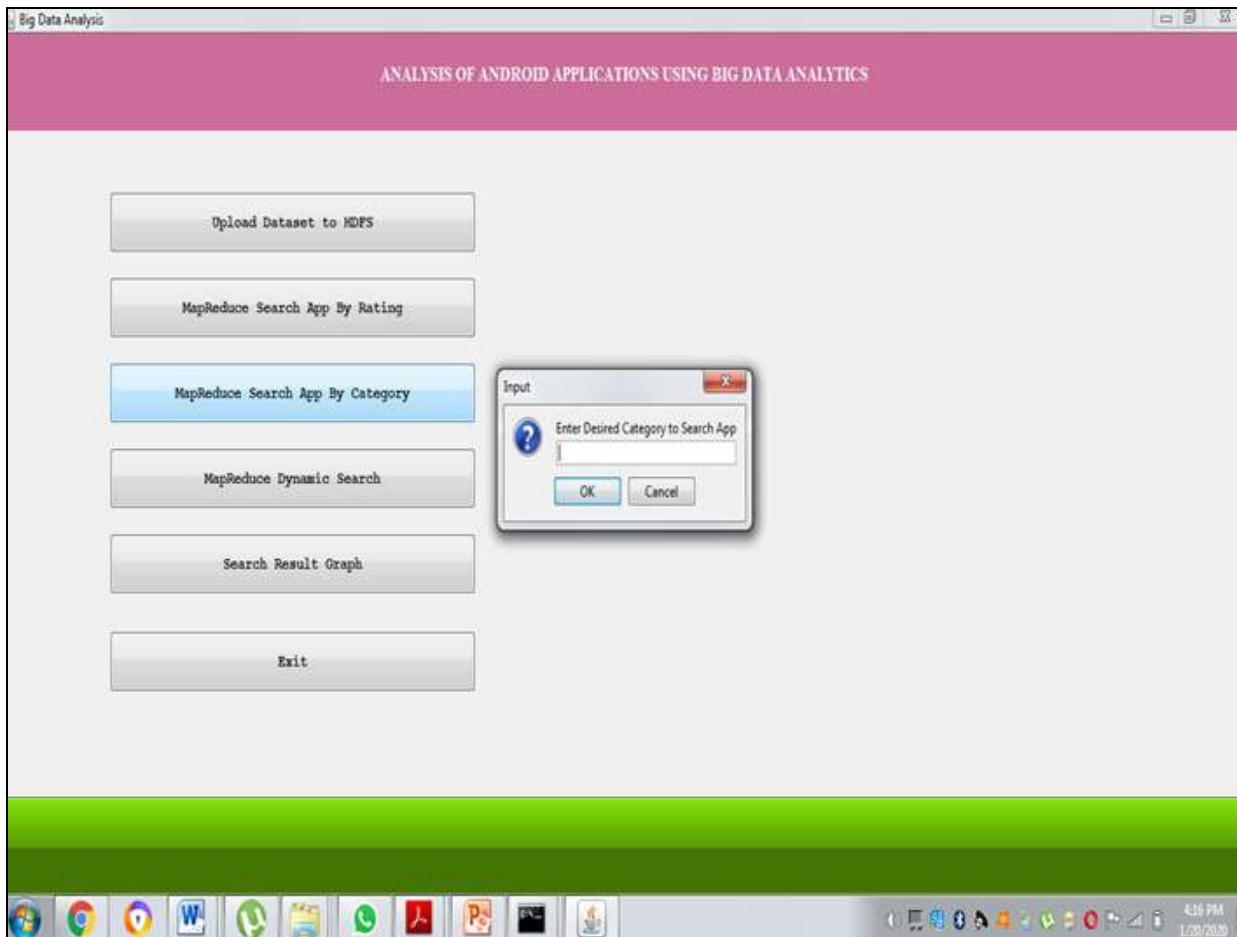


Fig 4: Map Reduce Search App By Category

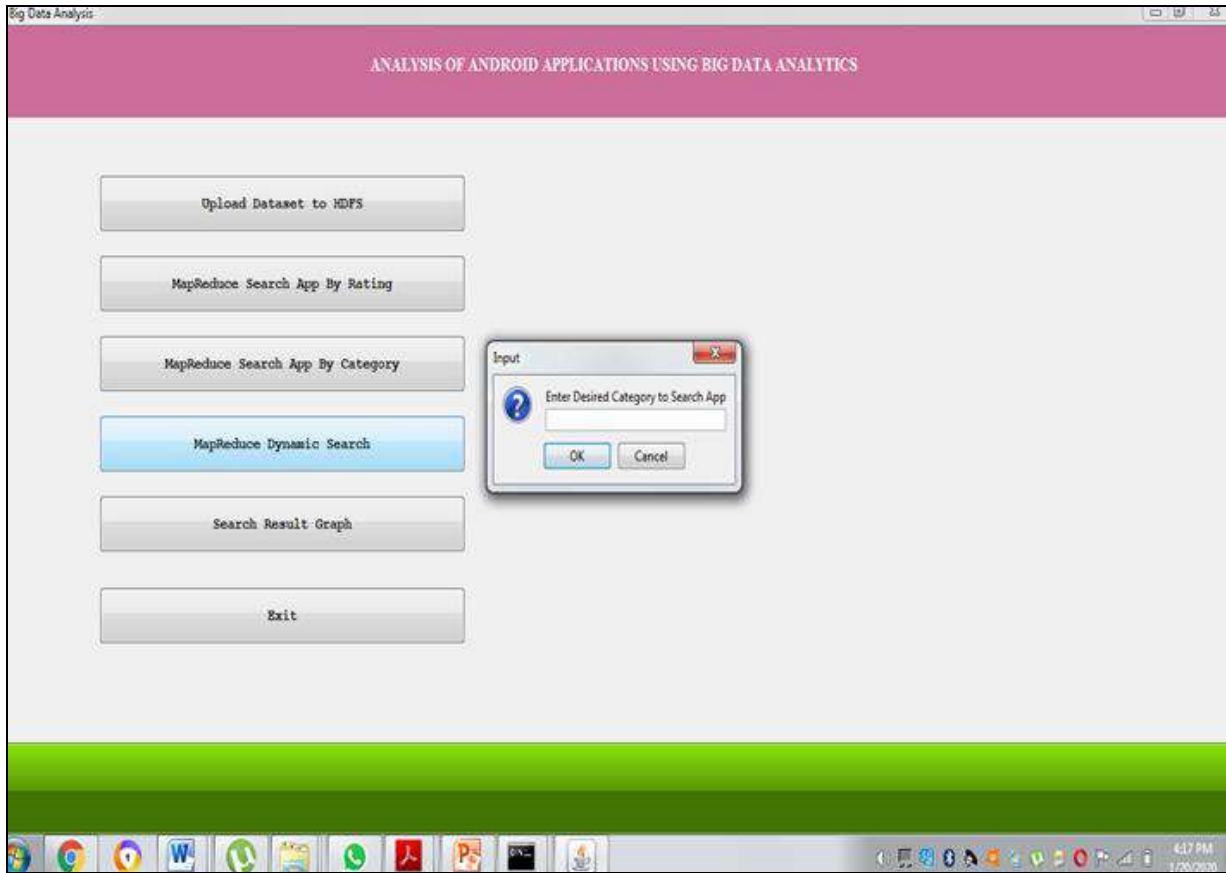


Fig 5: Enter Desired Category to Search App

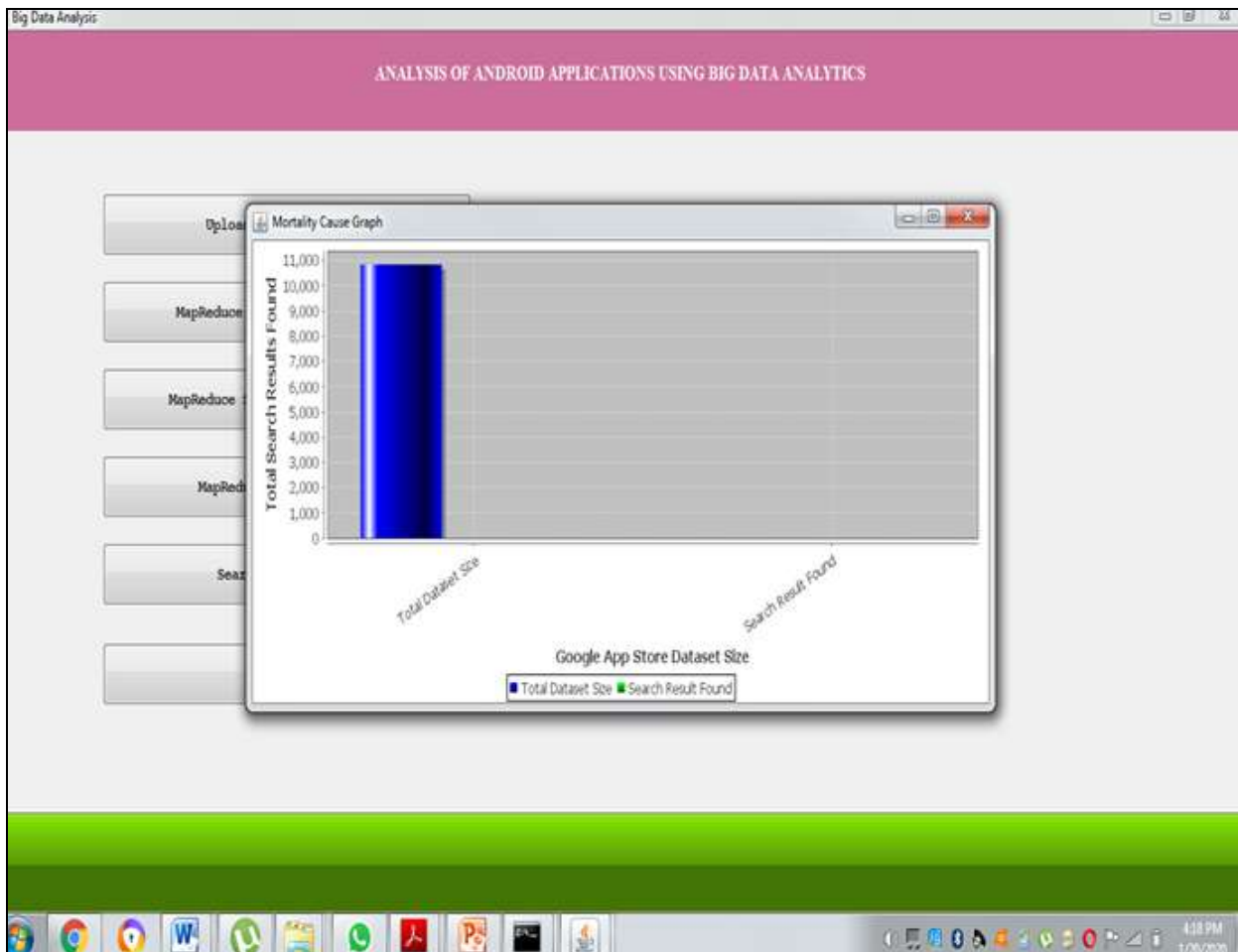


Fig 6: Graph Based On App's Data

4. Conclusion

Here, I have designed a project "ANALYSIS OF ANDROID APPLICATIONS USING BIG DATA ANALYTICS" using BIDATA. This is very beneficial for analysis Android App data. I presented my project which covers both proposed system and satisfies all the functional Requirements.

References

1. Brooks EG, Reed KD. "Principles and pitfalls: a guide to death certification," *Clinical medicine & research*. 2015; 13:74-82.
2. Jha P. "Counting the dead is one of the world's best investments to reduce premature mortality," *Hypothesis*. 2012; 10:e3.
3. Chorba TL, Holman RC, Clarke MJ, Evatt BL. "Effects of HIV infection on age and cause of death for persons with hemophilia A in the United States," *American journal of hematology*. 2001; (66):229-240.
4. Hooper WC, Holman RC, Clarke M, Chorba TL. "Trends in non-hodgkin lymphoma (NHL) and HIV associated NHL deaths in the United States," *American journal of hematology*. 2001; (66):159-166.
5. Mannino DM, Brown C, Giovino GA. "Obstructive lung disease deaths in the United States from 1979 through 1993: an analysis using multiple-cause mortality data," *American journal of respiratory and critical care medicine*, vol. 156, pp. 814-818, 1997.
6. C. Gordon, "Australian Bureau of Statistics, Multiple cause of death analysis. Publication 3319.0. 55.001," ed, 2003.
7. Bah S, Rahman MM. "Measures of multiple-cause mortality: a synthesis and a notational framework," *Genus*. 2009; (65):29-43.
8. W.H. Organization, International statistical classification of diseases and related health problems. World Health Organization, 2004, 1.
9. Israel RA, Rosenberg HM, Curtin LR. "Analytical potential for multiple cause-of-death data," *American journal of epidemiology*. 1986; (124):161-81.
10. Brown PF, Desouza PV, Mercer RL, Pietra VJD, Lai JC. "Class-based n-gram models of natural language," *Computational linguistics*. 1992; (18):467-479.