

Chapter 7

Prediction of Cancer Disease Using Classification Techniques in Map Reduce Programming Model

M. A. Saleem Durai
VIT University, India

Anbarasi M.
VIT University, India

Jaiti Handa
VIT University, India

ABSTRACT

As the volume of data is increasing with time the primary issue is how to store and process such data and get useful information out of it. Analysis of classification algorithms and MapReduce programming model has led to the conclusion that the distributed file system and parallel computing attributes of MapReduce are good for designing classifier model. The major reason for it is parallel processing of data in which data is divided and processed in parallel and the output from each is reduced further for a single output. In this paper, we are going to study how to use MapReduce model to build classifier model. We are using cancer dataset to predict if a person has cancer or not by using Naive Bayes and KNN classification algorithms. We have compared them on the basis on computational time and the factors like sensitivity, specificity, and accuracy. In the end, we would be able to compare these two algorithms and tell which one works better on MapReduce programming model

DOI: 10.4018/978-1-5225-2863-0.ch007

1. INTRODUCTION

“The amount of data that was generated during the last couple of years accounts to 95% of the total data” With the advancement of new technologies, increased usage of social networking sites, need of storing the data for analysis purpose is of utmost importance. Until yesterday the data you stored on the servers in your company was simply data, then suddenly a term emerges Big Data; this term refers to each and every bit of data you have stored till date. It includes even the URL’s you have been marking till date. In short, every piece of data doesn’t matter if it is structured or unstructured is collectively big data.

1.1 What Contributes to Big Data?

Data stored in Black Box: It is deployed in aircrafts. It keeps a record of all conversations of the crew, also records the information about the performance of the aircraft time to time. Data generated by social media: People on social networking sites share pictures, messages, voice messages, every second post has been generated and all such information needs to be stored. Stock Exchange: The stock market takes a new turn every minute, all the information regarding the “buying” and “selling” of shares needs to be tracked and stored. It further allows us to understand the market well.

1.2 Uses of Big Data

Understanding the Market and Customer Requirements: It is one of the biggest areas where big data is used. The market researchers will understand the market and help big firms and business act according to the trends in the market and the customer requirements for example it may help Wal-Mart to estimate which product to sell and at what cost according to the need and demand in the market.

Improve quality of healthcare service: Big data can be used to predict certain diseases like cancer, heart attacks etc. This will further assist our doctors to help provide quality of services (Ebenezer et al., 2015). It has the power to decode a DNA in seconds and help us to understand disease patterns.

Helps sportsperson to understand better and perform better: there are certain tools that use analytics using videos to understand how each player performs and to understand pattern each player follows. Such tools have already been used in analyzing performance of players in basketball, tennis.

Weather forecasting as well as disaster forecasting: Analytics has been used to forecast the weather conditions using big data. It is also used by scientists to predict disasters which further helps in saving human lives and prepare for such disasters in advance.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/prediction-of-cancer-disease-using-classification-techniques-in-map-reduce-programming-model/187663

Related Content

Forty Years of Federal Legislation in the Area of Data Protection and Information Security

John Cassini, B. Dawn Medlin and Adriana Romaniello (2011). *Pervasive Information Security and Privacy Developments: Trends and Advancements* (pp. 14-23). www.irma-international.org/chapter/forty-years-federal-legislation-area/45800

The Role of Knowledge Management on Job Satisfaction: A Systematic Framework

Kijpokin Kasemsap (2014). *Advances in Secure Computing, Internet Services, and Applications* (pp. 104-127). www.irma-international.org/chapter/the-role-of-knowledge-management-on-job-satisfaction/99454

Gender Dimension on the Living Conditions of Poor Household Heads in Sefene-Selam Sub-City, Bahir Dar, Ethiopia

Yetimwork Anteneh Wondim (2020). *International Journal of Risk and Contingency Management* (pp. 30-62). www.irma-international.org/article/gender-dimension-on-the-living-conditions-of-poor-household-heads-in-sefene-selam-sub-city-bahir-dar-ethiopia/261207

Effect of Cultural Diversity in Communication on Project Success Within Multinational Construction Companies in the United Arab Emirates

Brian J. Galli (2019). *International Journal of Risk and Contingency Management* (pp. 39-65). www.irma-international.org/article/effect-of-cultural-diversity-in-communication-on-project-success-within-multinational-construction-companies-in-the-united-arab-emirates/234433

Simulation Experiment of Key Exchange Protocol in Mobile Devices With E-Commerce Application

Pranav Vyas and Bhushan Trivedi (2020). *International Journal of Information Security and Privacy* (pp. 38-49).

www.irma-international.org/article/simulation-experiment-of-key-exchange-protocol-in-mobile-devices-with-e-commerce-application/256567