International Journal of Research in Advent Technology (IJRAT) Special Issue
E-ISSN: 2321-9637
Available online at www.ijrat.org
National Conference on "Role of Information Technology in Social Innovations"
26th & 27th February 2019

Subgraph Mining Algorithm on Big Data for Blood Cancer Detection on Big Data

Ashwini Abhale[1], Shreya Bramhankar[2], Prajkta Bijwe[3], Shabdali Sangale[4], Shwetambari Chaugule[5]
[1,2,3,4,5]Department of Information Technology D Y Patil College of Engineering, Akurdi

Abstract— Cancer is the leading cause of death worldwide. Therefore, identification of genetic as well as environmental factors is very important in developing novel methods of cancer prevention. However, this is a multi-layered problem. Therefore, a cancer risk prediction system is here proposed which is easy, cost effective and time saving. Health care systems generate a huge data collected from medical tests. Data mining is the computing process of discovering patterns in large data sets such as medical examinations. Blood diseases are not an exception; there are many test data can be collected from their patients**.**

1. INTRODUCTION

Cancer is one of the most common diseases in the world that results in majority of death. Cancer is caused byuncontrolled growth of cells in any of the tissues or parts of the body. Cancer may occur in any part of the body and may spread to several other parts. Only early detection of cancer at the benign stage and prevention from spreading to other parts in malignant stage could save a person's life. Treatment of blood cancers has undergone substantial improvements, resulting in increased rates of remission and survival. Remission occurs when there is no sign of cancer.

2. LITERATURE SURVEY

A. Survey of Fault Detection, Isolation, and Reconfiguration Methods.

Fault detection, isolation, and reconfiguration (FDIR) is an important and challenging problem in many engineering applications and continues to be an active area of research in the control community. This paper

presents a survey of the various model-based FDIR methods developed in the last decade.

B. *Application of Local Outlier Factor method and Back-Propagation Neural Network for steel plates fault diagnosis*

Fault diagnosis, which is a task to identify the nature of the occurred fault, is of paramount importance to ensure the steadiness of industrial and domestic machinery.

Essentially, fault diagnosis is a problem of classification. A method based on Local Outlier Factor (LOF) anomaly detection and BP neural network is proposed to apply to steel plates fault diagnosis.

3. FUTURE SCOPE

On the basic of image data system will detected blood cancer .In blood cancer there are many stages First, Second And Third It will detected in details. also generate models that can distinguish patients with normal blood disease from patients who have blood cancer. We evaluated our results using different technique applied on real data collected from hospitals.

A. Proposed System

System used dataset of blood cancer in back end. Load dataset on map reduced. Apply filter on the dataset used apriori algorithm. By providing the attributes as a input system generate the result whether the patient has cancer or not. Basically, it gives the result for early stage of cancer. By using the Hadoop platform it gives the parallel and faster processing of data.

B. SYSTEM SPECIFICATION

Hardware Requirements:

• System	:	Pentium IV 2.4 GHz.
• Hard Disk	:	40 GB.
• Floppy Drive	:	1.44 Mb.
• Monitor	:	15 VGA Color.
• Mouse	:	Logitech.
• Ram	:	512 Mb.

Software Requirements:

• Operating system	: Windows XP/7.
• Coding Language	: JAVA
• IDE	: Eclipse
• Database	: MYSQL

B. System Architecture

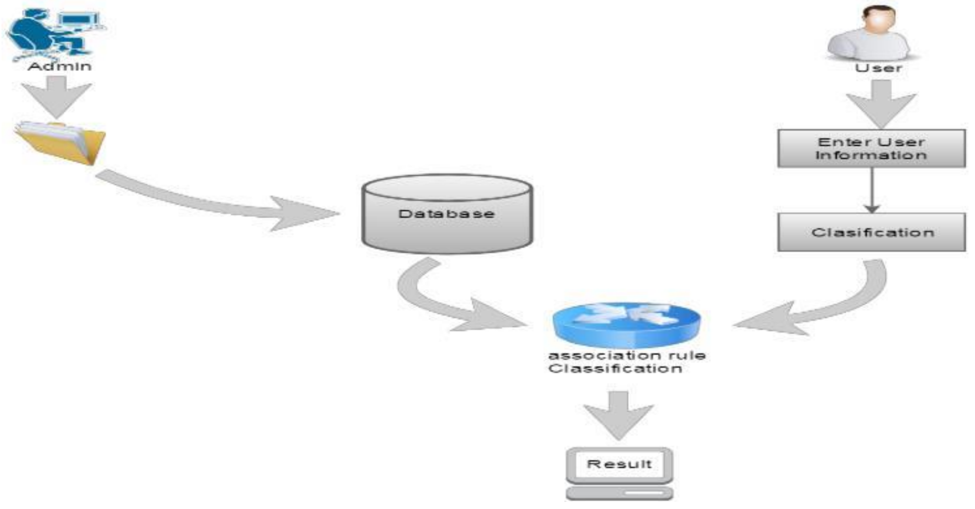


Figure: System Architecture of Proposed System

A. Algorithms

1. APRIORI ALGORITHM

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database this has applications in domains such as market basket analysis.

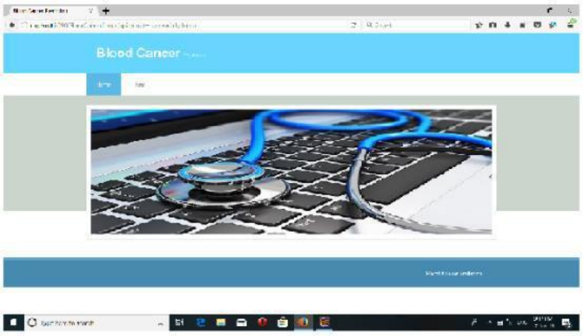
The Apriori algorithm was proposed by Agrawal and Srikant in 1994. Apriori is designed to operate on databases containing transactions. Other algorithms are designed for finding association rules in data having no transactions, or having no timestamps. Each transaction is seen as a set of items (an *item set*). Given a threshold the Apriori algorithm identifies the item sets which are subsets of at least transactions in the database. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as *candidate generation*), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

B. Modules.

- User Module
- Admin Module
- Prediction Module
- Final Result

1. User Module

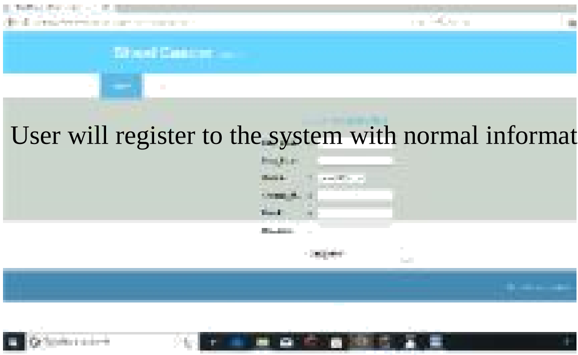
- Login



2. Admin Module

- **Registration**
- **Login**
- **Search for Data**

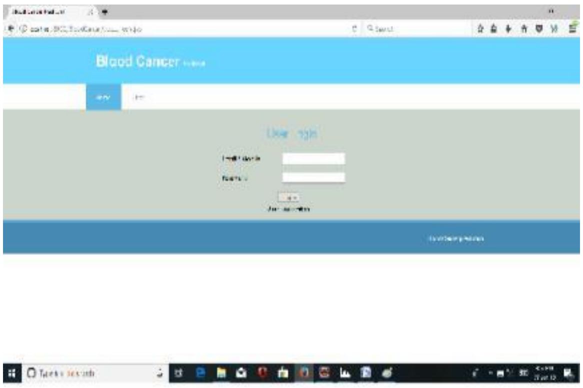
A. Registration



User will register to the system with normal information.

B. s Login

For login the user will enter the user name and password, if entered information is correct then the system will redirect to the home page, otherwise it will show an error message.



C. Search for Data

- After login the user will search for data which he/she required.
- Then user will get ranked data from structured database.
- Then he/she can download the file and check the result.
- check the result.

ACKNOWLEDGMENT

I have taken efforts in this project. However, it
would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them. I am highly indebted to Mrs. Ashwini Abhale for his guidance and constant supervision as well as for providing necessary information regarding the project & also for his support in completing the project. I would like to express my gratitude towards my parents &my Head of I.T.

Department for their kind co-operation and encouragement which helped me in completion of this project. Furthermore, I would also like to acknowledge with much appreciation the crucial role of the staff of DYPCOE Akurdi , who gave the permission to use all required equipment and the necessary materials to complete my project-based seminar. I also am deeply grateful to the Principal of DYPCOE, Mr. V. M. Wadia and my parents for their financial and logistical support and for providing necessary guidance concerning project's implementation.

REFERENCES

[1] Rosenthal, P. Mork, M. H. Li, J. Stanford, D. Koester, and P. Reynolds, ʊ Cloud computing: Cloud computing: A new business paradigm for biomedical information sharing, Journal of Biomedical Informatics, 2010.

[2] W. Lam, L. Liu, S. Prasad, A. Rajaraman, Z. Vacheri, and A. Doan, ʊ Cloud computing: Muppet: Map reduce-style processing of fast data, Proc. VLDB Endow, 2012.

[3] G. Liu, M. Zhang, and F. Yan, ʊ Cloud computing: Large-scale social network analysis based on map reduce, Intl. Conference on Computational Aspects of Social Networks (CASoN), 2010.

[4] J. Dean and S. Ghemawat, ʊ Cloud computing: MapReduce: simplified data processing on large clusters, Commune, ACM, 2008.

[5] U. Kang, C. E. Tsourakakis, and C. Faloutsos, ʊ Cloud computing: Pegasus: A peta-scale graph mining system implementation and observations, in Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, 2009.

[6] U. Kang, B. Meeder, and C. Faloutsos, ʊ Cloud computing: Spectral analysis for billion scale graphs: discoveries and implementation, ۏ in Proceedings of the 15th Pacific Asia conference on Advances in knowledge discovery and data mining - Volume Part II, ser. PAKDD'11, 2011, pp. 13–25.

[7] Suri and S. Vassil vitskii, ʊ Cloud computing: Counting triangles and the curse of the last reducer,
ۏ in Proceedings of the 20th
international conference on World wide
web, 2011.