ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

Random Forest Data mining approach to Determine Accuracy by using fibroid Dataset

Girija D.K
Research Scholar
MUIT, Lucknow.
girijadk16@gmail.com

Dr. Manish Varshney
Professor,
MUIT, Lucknow.
itsmanishvarshney@gmail.com

Abstract: The amount and sensitivity of data in healthcare is enormous. The data must be treated with extreme care, and no shortcuts should be taken. In demand to estimate the quality of healthcare services, a numeral of data mining classification methodologies have been utilized. This study discusses and assesses the experience of implementing a data mining approach and procedures on the basis of 150 patient records. This novel approach for determining the accuracy of a product has been developed using data mining technique. A variety of techniques, such as Decision Tree, Naive Bayes, KNN, Radom Tree Set, Rule Model, ZeroR and J48 or C 4.5 are there in data mining. Among this technique, I used the Random Forest data mining classification to predict the accuracy by using the fibroid data set. Our goal is to analyses the accuracy and the other indicators values like RMSE, Recall and Precision etc. in order to produce a conclusion.

Keywords: Data Mining, Healthcare, Accuracy, Techniques and Fibroid.

1. Introduction

Data mining has recently gained popularity due to the amount of data and the imminent need to extract information and knowledge from it. Data mining is a logical evolution of information technology. Data group and database construction, data administration (including data storage and recovery, and database operation processing), and new data analysis have all advanced in the database system company (including data warehousing in addition to data mining).

Because of its huge success, data mining is becoming increasingly popular in the healthcare industry. We used the fibroid data set. We choose it since it is a global disease. Female reproductive system fibroids are the most common type of tumour. Fibroids are uterine tumours made up of smooth muscle cells and fibrous connective tissue. This study analyses several categorization methods.

Data Mining: Analysis of huge data sets can help businesses address challenges by identifying patterns and linkages that can be discovered through **data mining**. These strategies and tools allow businesses to forecast future trends and make more informed decisions.

Types of Data Mining: Data mining is a method of obtaining knowledge from large databases. Types of data are **descriptive and predictive** (Hong and Weiss, 1999; Han and Kimber, 2002). The **descriptive** model is used to summarize data and develop conclusions. Descriptive data mining is used for database summary and visualization. Using several degrees of abstraction, one can investigate a data set's general behavior, which is difficult to derive from a huge database. In contrast, **predictive** analytics is focused on creating models that can predict future events. Two prominent mining tasks are cataloguing and regression.

Data mining requires classification. Indeed, data mining is a sort of machine learning influenced by pattern recognition, a science tasked with classifying things into a number of distinct categories, referred to as classes. Patterns are little data units that are specific to a situation.

2. Healthcare Application In Data Mining

There is simply too much data to collect and analyse using standard approaches when it comes to healthcare transactions. Identifying patterns and developments in vast amounts of composite data helps improve decision-making. The necessity for healthcare organisations to make decisions based on the investigation of medical and financial data has grown as financial pressures have increased. A high level of care can be maintained while using data mining insights to influence cost, revenue, and operating efficiency.

3. Scientific Relevance and Development

This study includes just a small number of patients. There may be a variety of radiation dosage decreases in a bigger research, both in terms of proportion and total. As a result, one of the 25 patients who participated in the research received further OAE, despite the fact that less than one percent of patients demonstrated that we had such large collateral flow utilising aorthographs. We feel that larger studies are more likely to correctly reflect the genuine occurrence. Aortograms after embolization were examined to see if they may alter holy arterial flow shapes and hence lower ovarian artery angiographical empathy. Having stated that, we feel that clinically important ovarian migration to fibroids is best discovered following the conclusion of the first supply from uterine arteries.

4. About Fibroid

A uterine fibroid is a benign (non-cancerous) growth that is common in women's uteruses (womb). Fibroids are tumours that form in the uterine hedge, which is where they

are most commonly encountered in women. They can either mature or reproduce within the uterine wall. It can spread alone or in groups. Obstetric fibroid tumours are classified as "benign" since they do not produce symptoms such as heavy menstrual flow or incontinence (not cancerous).

In women, fibroids are unusual growths in or on the uterus. Fibroids are benign growths of connective tissue and smooth muscle in the uterus. Tumors of this kind can develop to enormous proportions, resulting in excruciating abdominal discomfort and erratic sleeping patterns.. In some cases, they don't show any signs or symptoms at all.

There are two types of fibroids: those that form on the **inner and outside** walls of the uterus, and those that form in the middle. Many disorders can cause fibroids, including as tumours, myomas, and leiomyomas, however the most common is fibroids. Despite the fact that they are not malignant, they have been found to be the core cause of many problems. Most women don't know they have fibroids until they have an issue with them.

5. Causes of Fibroid

There is a great deal of reciprocity in the uterine fibroid. Uterine fibroids are a subject that many women have an opinion on throughout their lives. Most women's uterine fibroids are too tiny or undetectable to pose a threat. Oestrogen, a female hormone, appears to play a role in the development of uterine fibroids. Fibroids in the uterus do not develop until after puberty and then only on a sporadic basis until age 30. Because oestrogen levels begin to decline after menopause, fibroids tend to shrink or disappear. They are more likely to suffer from uterine fibroids symptoms, and African American ladies have a difficult risk of developing fibroids than white ladies. Doctors don't know much about uterine fibroids, but what little they do know points to the following risk factors,

- **Genetics:** Changes in the genetics many fibroids contain gene changes that are distinct from those observed in normal uterine muscle cells.
- **Hormones**: Fibroids appear to be stimulated by oestrogen and progesterone, two chemicals that promote the development of uterine covering throughout each menstrual cycle. Fibroids have greater levels of oestrogen and progesterone receptors than normal uterine muscle cells. When a woman reaches menopause, fibroids tend to fade away. In addition to these reasons, uterine fibroids may develop as a result of,
- **Pregnancy:** Fibroids are less common in women who have children.
- **Early menstruation:** A woman's risk of developing uterine fibroids increases if she had her first period before the age of 10.
- **Family history:** Having fibroids in the uterus is a side effect.

6. Uterine Fibroids Diagnosis

When a doctor does a physical pelvic examination, he or she is more likely to notice uterine fibroids of all sizes. To establish the presence of uterine fibroids, imaging exams are carried out on a regular basis.

- **Ultrasound:** Heavy-frequency sound waves were bounced off the uterus and pelvic edifices by an ultrasound probe put in the vagina or above the abdomen's pelvis. A cinematic depiction of the uterus and any uterine fibroids is depicted.
- **Magnetic resonance imaging (pelvic MRI):** Use of an MRI scanner generates incredibly detailed pictures of the uterus and other pelvic structures thanks to a strong magnet and a computer. If the diagnosis of uterine fibroids is unquestionable, a pelvic MRI can be used to confirm their presence.

7. Statement of the Issue

- Problem While Indian healthcare has long been a source of national pride, there is room to improve patient outcomes and reduce healthcare costs. In modern healthcare, blocking measures are required. More and more studies show that uterine fibroid medicines work.
- Uterine fibroid tumours cause excessive bleeding that has killed numerous women worldwide. In rural locations, where innocent girls could not gain such knowledge, there was only great suffering and death. No quick access to data or symptoms means no low-cost diagnosis.

8. What is the study's goal?

- Data mining techniques and based systems in the Indian healthcare system will have a significant impact on all levels of the 0–1 scale in the future.
- By applying the Random Forest classifier, calculation of the accuracy and other factors by using weka data mining tool to our dataset

9. About Weka

The Weka workspace is a collection of machine learning and data preparation methods. Data input, statistical evaluation of learning outlines, and visual representation of input and learning results are all included in this comprehensive framework. Using flexible methodologies and new datasets, it is designed to quickly stab out existing approaches. A wide range of preprocessing techniques are included in a wide variety of learning methods. Using this tools, users may evaluate and contrast multiple approaches to find the ones that are most effective for their particular set of challenges.

Weka is an acronym for "Waikato Atmosphere for Knowledge Analysis," the name given to the structure developed at University of Waikato in New Zealand. The Weka, a non-flying bird native to the islands of New Zealand with a nosy disposition (its name rhymes with "Mecca"), may be spotted outside the university grounds. The GNU General Public License governs the use and distribution of the programme, which is written in Java. It's been tested on Linux, Windows, Macintosh, and even a personal digital assistant under these operating systems. It offers a consistent boundary to a wide range of learning algorithms, as well as approaches

for pre- and post-processing and calculating the results of education structures on any given dataset.

Data Set Description	
Attribute Name	Description
Age	Age
STATUS	status(Married, Single)
HB	Heavy Bleeding(3 to 4 days - No, More than 7 day-HIGH)
PP	Pelive Pan(High, No)
FT	Fibroid Type(INTRACAVITIARY,SUBMUCOSAL,SUBSEROSAL,PEDUNCULATED,INTRAMURAL)
LBP	Lower Backpain(High, NO)
PDI	Pain During intercouse(High, NO)
FU	Frequnt Urination(Yes, No)
NFP	Number of Fibroid Present(Multiple, Single)
SF	size of fibroid(1mm to 20CM-(8 inches) in diameter or even larger)
CAUSES	Causes(INFERTILITY ,ANEMIA ,SWELLING IN THE ABDOME, NO EFFECT OF FERTILITY,PREVEMENT SPERM, NO EFFECT,EFFECT)
CLASS	Class(Eliminate, KEEP)

Table 1: Illustrations the different features used in our approach

The above table: 1 illustrations the different features and their descriptions that that should use in this paper for implementation purpose, this data are collected from reputable hospitals have been gathered and analysed. Due to the fact that additional data is confidential and cannot be disclosed, the study restricts access to 12 fields, which contain the following attribute names: age, status, HB, PP, FT, LBP, PDI, FU, NEP, SF, CAUSES, and CLASS.

10. Weka Implementation Using Random Forest Classifier

RANDOM FOREST: There are several applications for the supervised data mining approach known as random forest. However, the majority of the time, it is used to solve classification problems. Trees make up a forest, and a forest with more trees is a healthier forest. Additionally, the random forest technique builds decision trees from data samples, gathers predictions from each, and then votes on the best alternative. As a result, it is better than a single decision tree in terms of reducing overfitting.

Functioning of Random Forest Algorithm

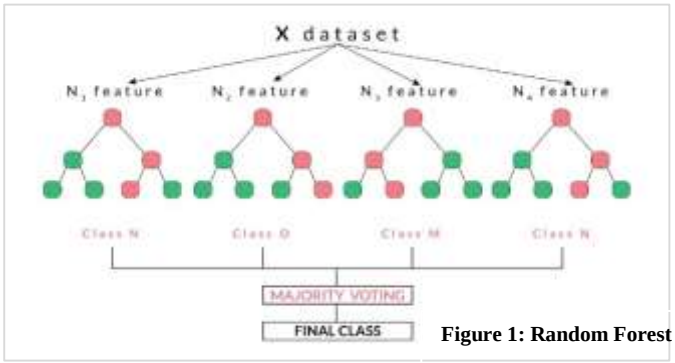

Figure 1: Random Forest

Step 1: Random forest selects n random records from a record set of k records.

Step 2: For each example, an distinct decision tree is made.

Step 3: The output of every decision tree is generated.

Step 4: Averaging or Majority Voting are used to determine the final classification and regression results.

Random Forest Algorithm Pseudo code

To produce c classifiers:

In the range of i=1 to c do

Randomly sample the training data D with replacement to produce D_i

Create a root node, N_i containing D_i

Call Build Tree(N_i)

end for

Build N Tree;

if N contained occurrences of first unique class then

Return

else

Randomly choice x% of the possible unbearable features in N

Select the feature F with the highest information gain to split on.

Create f child node of N- N_1, N_2……N_f, where F has f possible values (F_1, …… F_n)

In the range of i=1 to f do

set the fillings of Ni to D_i, where D_i is all occurrences in N that equal

Fi

Call BulidTree (N_i)

End for

End if

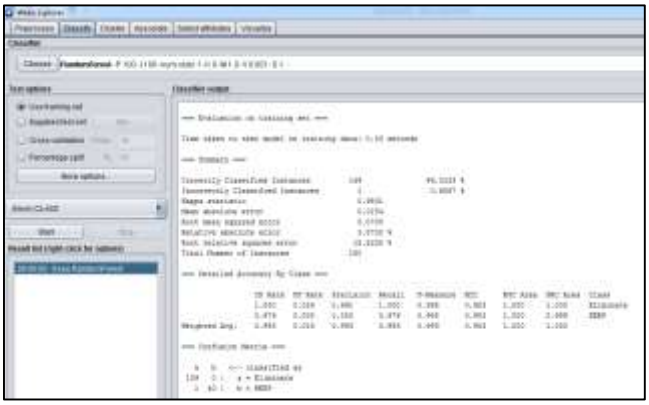

Figure 2: Random Forest Classifier Result

Different Indicators of the Random Forest Classifier

Classifiers /Indicators	Accuracy	RMSE	ROC Area	Precision	Recall	F-measure
Random Forest	99.33%	0.0705	1	0.991	1	0.995

11. Conclusion

Using random forest the accuracy number for classifier is displayed. In order to save time in predicting and treating patients, datasets in the labelled class-eliminate might be disregarded. The "classes-keep" dataset, on the other hand, is utilised in the medical and instrumentation areas to predict future treatment results.

Accuracy of the Random Forest classification algorithm is calculated and it is represented in the above figure, from the results **Random Forest** algorithm classifies the given data 99.3333% correctly, RMSE = 0.0705, Recall = 1.000 and it is useful to decide whether we have to **keep** the fibroid patient in the hospital for treatment or **eliminate** if it is starting stage.

Reference

[1]. Theodore Dalamagas, PanagiotisBouros, Theodore Galanis, MagdaliniEirinaki, and TimosSellis. Mining user navigation patterns for personalizing topic directories. In 9th InternationalWorkshop on Web Information and Data Management, pages 81-88, 2007.

[2]. Peter Brusilovsky. “Methods and techniques of adaptive hypermedia". User Modeling and User-Adapted Interaction, 6(2):87-129, 1996.

[3]. BamshadMobasher, Honghua Dai, Tao Luo, and Miki Nakagawa.Improving the effectivenessof collaborative filtering on anonymous web usage data.Technical report, 2001.

[4]. Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in informationsystems research. MIS Quarterly, 28(1):75-105, 2004.

[5]. Alberto Pan, Juan Raposo, Manuel Alvarez, Paula Montoto, VincenteOrjales, Justo Hidalgo,Lucia Ardao, AnastasioMolano, and Angel Vina. The denodo data integration platform. In 28thInternational Conference on Very Large Data Bases, pages 986-989, 2002.

[6]. AykutFirat. Information Integration Using Contextual Knowledge and Ontology Merging.PhDthesis, Massachusetts Institute of Technology, 2003.

[7]. Ricardo Baeza-Yates and Berthier A. Ribeiro-Neto. Modern Information Retrieval.ACM Press,1999.

[8]. Sanjay K. Madria, Sourav S. Bhowmick, Wee Keong Ng, and Ee-Peng Lim. Research issuesin web data mining. In 1st International Conference on Data Warehousing and KnowledgeDiscovery, pages 303-312, 1999.

[9]. Jose Borges and Mark Levene.Data mining of user navigation patterns. In Workshop on WebUsage Analysis and User Pro_ling, pages 31-36, 1999.

[10]. DominikFlejter, Karol Wieloch, and WitoldAbramowicz.Unsupervised methods of topical textsegmentation for polish. In Workshop on Balto-Slavonic Natural Language Processing, pages51-58, 2007.

[11]. Susan Dumais, John Platt, David Heckerman, and MehranSahami.Inductive learning algorithmsand representations for text categorization. In 7th International Conference on Informationand Knowledge Management, pages 148-155, 1998.

[12]. Michael Steinbach, George Karypis, and Vipin Kumar.A comparison of document

[13]. Lawrence Kai Shih and David R. Karger. Using URLs and table layout for web classificationtasks. In Stuart Feldman, Mike Uretsky, Marc Najork, and Craig Wills, editors, 13th InternationalConference on World Wide Web, pages 193-202, 2004.

[14]. https://www.upgrad.com/blog/data-mining-techniques/