RESEARCH ARTICLE

SAMRIDDHI Volume 15, Issue 2, 2023 Print ISSN: 2229-7111 Online ISSN: 2454-5767

Web Mining and Anemia Classification in Gujarat via WWW

Kajal Thakor[1], Swapnil Parikh[2], Vipul Dabhi[3], Ankita Gandhi[4]

Computer Science and Engineering, Parul Institute of Engineering and Technology, Vadodara, Gujarat, India.

Abstract

Web mining makes extracting internet data from big databases feasible. Low hemoglobin or insufficient red blood cells are symptoms of anemia. Undernutrition is the most common cause, and it is most common in young children, pregnant women, and women of childbearing age. If it is not treated or doesn't take into consideration, it could lead to an enlarged heart or heart failure. In this study, we gather data from the web about anemia in Gujarat and categorize it in various ways based on its parameter. The dataset builds from web scrapping using Python soup. The dataset we extract can be further classified to make a decision. This research helps in identifying anemia-related information as well as saving lives. Data mining tools predict future trends, help organizations, and present knowledge in the form of the easiest way so that humans can understand. The data can be processed, analyzed, and presented by a specifically trained algorithm for diverse stakeholders, including policymakers, program planners, health service providers, academicians, research scholars, and common citizens, enabling them for evidence-based, data-driven decision-making.

Keywords: Web mining, Web scraping, Data mining, Anemia, Data science, Machine learning, Classification, Data extraction, Worldwide, and web.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2023); DOI: 10.18090/samriddhi.v15i02.09

Introduction

Data mining techniques are used in the practice of "web mining," which involves automatically identifying and extracting information from websites. Figure 1 shows how web mining operates. Finding relevant information from the World Wide Web and its patterns is the fundamental goal of web mining. Blood, saliva, and water are just a few of the liquids that make up the human body. In which the maintenance of hemoglobin and system regulation are crucial functions of blood. A lower-than-normal blood volume brings on anemia. One of the most prevalent diseases in the world is known by the word "anemia."It has received a lot of attention because it is an international public health problem usually affecting young children, expectant mothers, and the elderly.

Therefore, anemic sufferers may have shortness of breath, pallor, cold hands, dizziness, reduced immunity, and palpitations. If anemia symptoms are not treated, they will worsen over time or potentially result in death. Anemia symptoms might worsen over time or potentially result in death if they are not treated.

What is Web Mining, exactly? What are its varieties?

Web mining is the practice of using data mining techniques to harvest details about online records and services, server logs, and hyperlinks in order to spot and gather interesting patterns, trends, and data in general (Figure 2). Web mining's primary objective is to discover patterns in web data by gathering and analyzing data to gain significant insights. The categories for web mining are displayed in Figure 3.

There are several varieties of web mining, including the ones listed below:

Web content mining is a Web mining technique that involves extracting crucial descriptive data from websites (WWW). Audio, video, text documents, hyperlinks, and structured records all constitute content. Text, lists, photos, videos, and tables are just a few of the content types used on websites to present records to consumers.

Data extraction, which involves copying structured data from unstructured websites, is the purpose of content mining. The retrieved structured data will be used to assist data aggregation across several websites.

Corresponding Author: Kajal Thakor, Computer Science and Engineering, Parul Institute of Engineering and Technology Vadodara, Gujarat, India, e-mail: thakorkajal6@gmail.com

How to cite this article: Thakor, K., Parikh, S., Dabhi, V., Gandhi, A. (2023). Web Mining and Anemia Classification in Gujarat via WWW. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 15(2), 241-248.

Source of support: Nil

Conflict of interest: None

Figure 1: Web mining

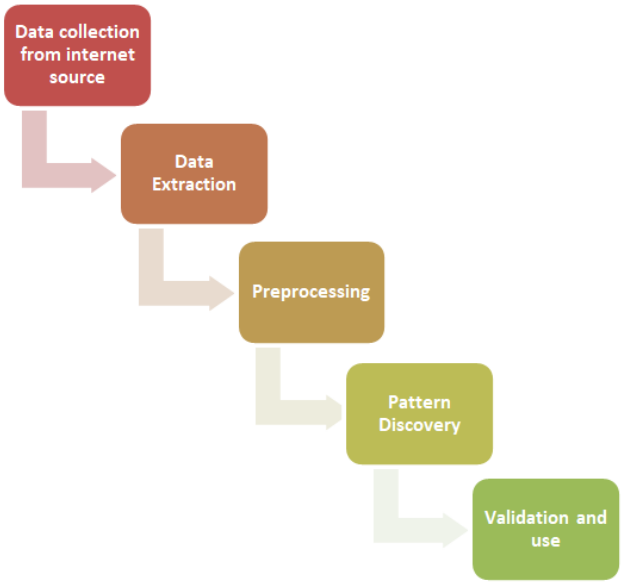

Figure 2: Depicts schematic overview of the processes involved in web mining

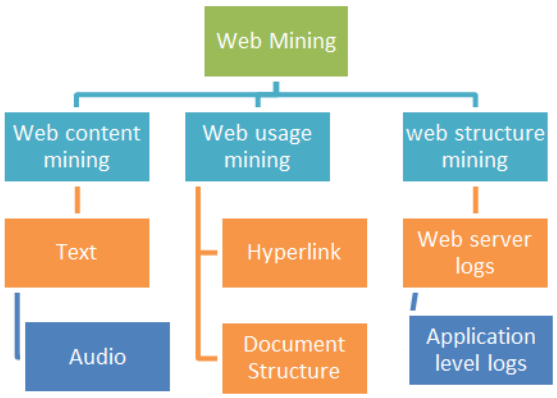

Figure 3: Web mining categories

One primary web mining approach that works with hyperlink structure is called web-structured mining. A structured overview of the website is revealed through structure mining. It can identify connections among websites' linked web pages.

Data mining that simply includes information retrieved from the web is known as web mining. The web is searched for data using a variety of algorithmic techniques. Structure mining analyses the website's hyperlinks to compile educational records and categorise them in terms of similarities and relationships. Implemented at the document level, intra-page mining is a subset of inter-page mining, which is mining at the hyperlink level.

Web usage mining is used to elicit pertinent data, expertise, and records from blog data and assists in determining how often users view particular home pages.

When someone uses web resources for mining, they are thinking about the logs of visitor requests to websites, which are frequently kept as web server logs. While the structure and content of the collection of home pages reflect the writers' goals, individual requests reveal how visitors interact with the pages. Relationships that the page's author had not intended to be revealed may be found through web usage mining.

Why is Web mining necessary?

Finding usable information from the World Wide Web and its usage patterns is the fundamental goal of web mining. Making use of web mining:

- By categorizing web content and locating web pages, web mining increases the effectiveness of web search engines.
- It is utilized for vertical search (such as Fat Lens, Become, etc.) as well as web search (such as Google, Yahoo, etc.).
- User behavior is predicted via web mining.
- Web mining is particularly helpful for a specific website and e-service, such as landing page optimization.

Related Works

Dogan *et al.*[1] describe Data Mining, Pattern recognition, Feature extraction, and Hematology Parameters used for the detection of anemia. The study created a decision support system that substantially facilitates and aids doctors in identifying iron-deficient anemia. Using data mining techniques, a decision tree performs system activities based on the pattern recognition process. Serum iron, serum iron-binding capacity, and ferritin characteristics are the fundamental hematological parameters that are employed to enter the system, with final anemia (+) and anemia (-) findings being justified at the end of the process.

S. J. Mohammed MOHAMMED *et al.*[2] mark out anemia prediction based on the classification rule with the help of Data Mining, ZeroR algorithm, OneR, and PART algorithm. Data are gathered for each of the 10 pertinent features in this paper. The development of a curated anemia prediction system involves the use of three classification algorithms.

Relevant anemia datasets are extracted by ZeroR, OneR, and PART using the "if" and "then" methodology. Based on the methodology used, PART outperformed ZeroR and OneR in accuracy by a factor of 85%. These methods offered guidelines for additional methods that were used to clarify the accepted understanding of anemic data norms.

Abdullah *et al*.[3] discuss Data mining, classification algorithms, naïve Bayes, J48 decision tree, Support vector machine, and SMO

for the prediction of anemia based on classification. In this study, we use a predictive model that maintains many data mining classification algorithms to categorize the anemia type of anemic patients. The complete blood count (CBC) test results from patients serve as the foundation for the dataset's actual data. We used WEKA's data mining tools to perform classification algorithms such Naive Bayes, Multilayer Perception, J48, and SMO after filtering this data to remove extraneous variables. It has been demonstrated through a number of experiments that the J48 decision tree algorithm offers the most accurate classification of anemia kinds.

Sasikala *et al*.[4] describe Data Mining, Classification Technique, J48, Random Forest tree, and Hoeffding Tree for predicting anemia. Data are collected in this survey. The open-source WEKA programme is used for research purposes. Data mining methods like Hoeffding trees, J48, and random forest trees are used in experimental studies. Finally, the accuracy and performance of the three categorization algorithms are compared. Compared to J48 and Hoeffding tree approaches, random forest trees are thought to be more accurate.

A. H. Shurrab *et al*.[5] outline three methods Decision tree, Naïve Bayes and Rule Induction. We use three data mining classifiers on test blood data in this study. The classifier makes early predictions about potential blood diseases using the CBC properties. For powers of healing, this is simple. Three separate hematological illnesses are the subject of three experiments. Oncology, Adult Haematology, and PediatricHematology. The findings demonstrate that the naïve Bayesian classifier is capable of foretelling malignancies in hematologic disorders. The decision tree performs the lowest when predicting hematology, while the rule-induced classifier performs better (57%–67%).

P. T. Dalvi *et al*.[6] use ensemble learning for the detection of anemia. This paper applies five ensemble learning methods to four classifier methods.The objective is to identify the classifier that classifies red blood cells most accurately to detect anemia. The consequence shows that the stacking ensemble strategy produces more accuracy than the other ensemble methods. The Artificial Neural Network and K-Nearest Neighbor perform the best and the worst among the individual classifiers, respectively. However, when used in a stacking ensemble, the Decision Tree and K-Nearest Neighbor classifier produce substantially higher accuracy than the Artificial Neural Network. As a result, an ensemble of classifiers outperforms a single classifier in terms of accuracy. So it's simple to get the highest level of accuracy while making medical decisions.

Laengsri V. *et al*.[7]describe how a discriminant model was built using machine learning techniques such as k-nearest neighbor (k-NN), decision trees, random forests (RF), artificial neural networks (ANN), and support vector machines (SVM). The observation was evaluated, correlated to 13 other discriminant formulas, and justified. ThalPred outperforms existing metrics and formulas, as shown by the predictive results of both cross-validation tests and independent data validation testing.

Sanap *et al*.[8] chose two techniques To determine the most effective method for anemia prediction and classification, C4.5 decision tree algorithm and support vector machine were used. Support vector machines and the C4.5 decision tree technique for classification are implemented in Weka as J48 and SMO in this work. Based on the CBC report and the anemia's severity, an Anemia Classification Determination Sheet is created to offer the most accurate anemia classification. The C4.5 algorithm, therefore, had the best performance and accuracy.

M. S. MOHAMMED *et al*.[9] specify Data Mining for the analysis of anemia. Based on available data, anemia has been predicted using four techniques: Bayesian Network (BN), Naive Bayes (NB), Logistic Regression (LR), and Multilayer Perceptron (MLP). In comparison to the other three strategies, the LR has produced better results. Additionally, attribute evaluators like information gain were used to show the system's good performance with few properties.

Dinakaran *et al*.[10] describe Data Mining, Data Pre-Processing, Normalization, and Clustering Analysis to find blood-related diseases. An innovative method for locating blood-related illnesses is created in this study. Complete blood counts are refined using data transformations such as normalization and data logging. It is reasonable to assume that the best results from enhanced data will be beneficial to medicine and human aid.

SrdjanDenic *et al*.[11] discuss Statistics, Setting, and Study Population for detecting Iron Deficiency and α-Thalassemia. The cumulative risk of all factors that reduced HbA2 is larger in this study than it was previously apparent. Due to the similarities in the characteristics of iron deficiency and + thalassemia and the fact that both conditions have low HbA2 levels, some BTT carriers may not be detected during screening. Women only; serum ferritin levels should be checked often.

F. A. Nugroho *et al*.[12] describe the causal discovery model, causality, iron deficiency anemia, iron supplementation, inflammation, and gut microbiome used for iron deficiency in children. Researcher predictions are assisted by computer models, which help to concentrate on causation and the most important aspects of the data. In order to determine the effectiveness of iron supplements and their connection to RTI, this study used a computer model markers of systemic inflammation and gut microbiota composition.

C. C. Hortinela *et al.*[13] describe raspberry pi, image processing, and support vector machine for the Identification of Abnormal Red Blood Cells. The suggested system measures a number of RBC properties. Similar investigations have already been carried out using a different methodology. One study classified RBCs with an accuracy of 90.54% using an artificial neural network (ANN), and another study classified RBCs with an accuracy of 83.3% using a radial basis function network. The proposed method, which distinguished between seven different red blood cells with 93.33% accuracy, utilized a support vector machine (SVM) classifier.

T. İlaslaner*et al.*[14] discusses Data Mining, Blood Biochemistry Parameters used for the deficiency of anemia. The planned system collected data from 50 sick and 50 healthy people. Artificial neural networks allowed for the development of the most successful data mining technique (78.31). Additionally, biochemical markers are useful in identifying iron deficiency anemia and aiding doctors in starting a patient's effective treatment.

Meena K *et al.*[15] describe three methods Data Mining, Decision Tree, and Associative Classification, used for the analysis of anemia. A healthcare model objective of minimizing the risk of anemia in blood-related diseases was suggested by the application of two techniques, decision tree and association rule mining, and the differentiation between them to choose a more suitable methodology for this specific work. The final decision tree's accuracy in predicting anemia is 97.35%.

Noor *et al.*[16] discuss Image Processing, Decision Tree, SVM, and KNN for Predicting Anaemic conditions.A total of 104 participants (male and female) will have their clinical blood hemoglobin levels, anemic state, and obtained photographs of the palpebral conjunctiva recorded for this study. A high-resolution mobile phone camera is used to shoot pictures. Her MATLAB image processing technique extracts the proportions of red, green, and blue pixels from the image. Hemoglobin levels are plotted using these characteristics. There are 81 total data collected, 23 of which

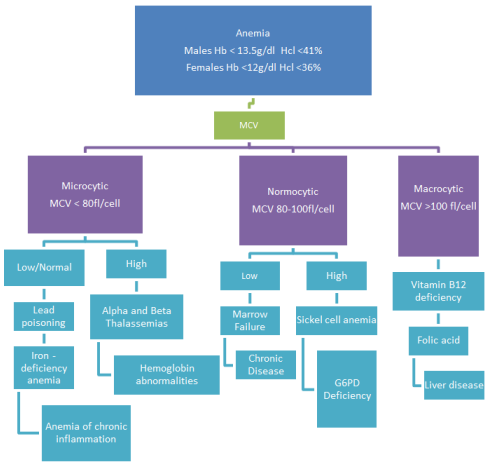

Figure 4: specific forms and/or causes of anemia

are for testing. To detect anemia, 81 data points were utilised to train several classifiers, including linear SVM, coarse trees, cosine KNN, etc.. Of the 23 data points tested, the decision tree produced the highest accuracy (82.61%).

K.Y. *et al.*[17] describe SVM, KNN, and Logistic regression for detecting anemia. In order to save time and effort in sickle cell pain management systems, this research focuses on detecting sickle cell disease with greater accuracy utilising various classifiers and developing better prediction models. The findings demonstrate that classification algorithms including SVM, ANN, logistic regression, decision trees, and random forests offer predicted accuracies of 76, 77, 78, 95, and 96%. The study comes to the conclusion that the random forest algorithm offers the highest level of predictive accuracy for detecting sickle cells in blood.

Methodology

The goal of the ongoing project is to create a system that will extract information on anemia and categorize it. In other words, we wanted to transfer the process of detecting different types of anemia, as the most prevalent condition, to the computer programme in order to teach it the decision-making process of an expert medical consultant. We gather all the data from web mining to accomplish this goal. Following the process of learning, we evaluated the system, classified it according to various criteria, and looked at its capacity to make decisions in a manner similar to that of an expert medical adviser. We applied the categorization procedure at this point. The study's primary finding relates to how web mining makes decisions. Another result is giving medical professionals, medical students, NGOs, and the WHO decision assistance. The expert first assesses the HGB value to check if anemia is present, as shown in Figure 4. The MCV value is checked in the subsequent step. The anemia type is microcytic if the MCV is 80. The MCV indicates the anemia type as normocytic if it is between 80 and 100. The anemia type is referred to be macrocytic if the MCV is higher than 100. The expert could need additional research and cutting-edge diagnostic procedures following the initial identification phase. Figure 4 details the specific forms and/or causes of anemia.

Actual patient data and the counsel of an experienced medical specialist are required for the computer to detect anemia like a knowledgeable medical consultant. This expert gave details on the criteria to meet and procedures to carry out when diagnosing anemia. For the method described in this study, each website needs more than one web page. This study includes websites from a wide range of areas, such as research, Bing, WHO, NGO, health care, academia, and google scholar. Web page extraction patterns were established for the website after page downloads. It is important to avoid making any qualitative alterations to the data because the goal is to make decisions like an expert medical specialist would make them. The outline schematic of the Web Scraping methodology is shown in Figure 5.

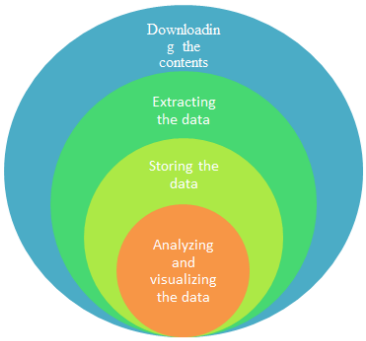

Figure 5: Outline schematic of Web Scrapping

Four fundamental models are created for the classification process when the data are collected via web scraping: support vector machines (SVM), decision trees (DT), artificial neural networks (ANNs), and Nave Bayes.

Types of anemia and diagnostic standards

The most prevalent group of blood diseases worldwide is anemia.Red blood cell numbers that fall short of what the body needs to function physiologically requirements are known as anemia, in compliance with WHO. A diminution in erythrocyte mass, blood hemoglobin concentration, or hematocrit is another definition of anemia. Age and gender affect normal hemoglobin and hematocrit readings. When hemoglobin and hematocrit levels are less than the upper bound of normal ranges for age and sex, anemia is present.

A decline in the morphological and/or physiological functioning of the erythrocytes is one of the main causes of anemia. There are four main causes of anemia: 1. an abnormality of erythrocyte production a. Misfunction or failure of the bone marrow b. Erythropoietin production is impaired; 90% of this substance is secreted from the kidneys and is crucial for the maturation of erythrocytes. 2. Impaired structural and functional erythrocyte maturation Early erythrocyte degeneration b. Erythrocyte degeneration causes (e.g., erythrocyte membrane disorders) 3. Causes of non-erythrocytes 4. Blood loss, weakness, weariness, and paleness are typical clinical signs of anemia. Leukemia and various other hematological disorders can cause bone and joint discomfort, swollen lymph nodes, and enlarged liver and spleen. Possible symptoms include heart palpitations, headaches, frequent infections, brittle nails, lack of appetite, loss of taste, painful tongue, mouth ulcers, and the desire to eat things other than food, such as ice, cement, or soil. Long-term anemic patients can cope with its symptoms more easily and may not have any serious concerns.

The effectiveness of a well-known online scraping technique is used in the current study to extract data. Techniques including ANNs, SVMs, decision trees, and Naive Bayes are used for classification. These cutting-edge classification techniques produce encouraging results. Additionally, it has been established that these techniques work well when applied to medical data. As a result, these techniques are used, and the outcomes are contrasted. If someone tries to scrape nonpublic data, they are breaking the law. Nonpublic data is not accessible to all people; attempting to extract such data is against the law.

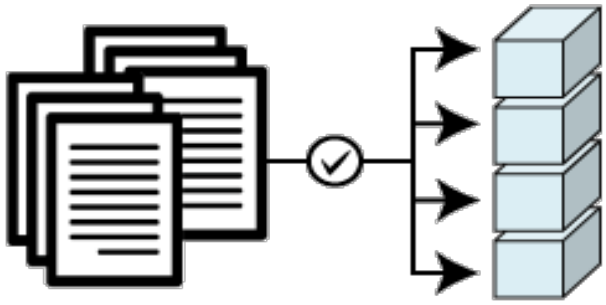

Figure 6: Web scraping working

The process of web scraping

A web scraper is a specialised programme made to swiftly and efficiently harvest data from numerous websites. Figure 7 shows how web scraping operates. Depending on the projects, web scrapers come in a wide range of designs and levels of complexity (Figure 6).

The steps to accomplish web scraping are as follows

Step 1: Locate the URL you wish to scrape.
Step 2: Examining the page
Step 3: Compose the code.
Step 4: Bookmark the information in the file.

Table 1: Icon used in this investigation and their meaning are listed below

ROC	Region of meeting the curve
SMO	Support vector machine
k-NN	K-nearest neighbor
DT	Decision tree
RF	Random forest
ANN	Artificial neural network
SVM	Supports vector machine
BTT	β -thalassemia trait
RBC	Red blood cell
MCV	Mean corpuscular volume
HB	Hemoglobin
RDW	Red blood cell distribution width
MCC	Matthew correlation coefficient
AUC	Area under the curve
BN	Bayesian Network
LR	Logistic Regression
MLP	Multilayer Perceptron
PCA	Principal component analysis
CFS	Correlation feature sub-set selection
NB	Naïve Bayes
IDA	Iron deficiency anemia

Table 2: Datasets Description

Method	*Data*	*Types/ diseases*	*Parameters*	*Accuracy*
Decision Support System [1]	96 patients (Hospital data)	Iron-Deficiency diagnosis	Serum iron, Serum iron-binding capacity, and Ferritin	(+) 76 Anemia pos. (-) 20 Anemia neg.
ZeroR, OneR and PART [2]	539 participants	6 Anemia types diagnosis	Mean absolute error, Root mean square error	PART-85%
NB, MLP, J48 DT, SMO [3]	41 patients	All types of Anemia diagnosis	TP Rate, FP Rate, Precision, F-Measure, Recall	J48 decision tree- ROC0.97
J48, Random Forest tree, Hoeffding tree [4]	200 households of students	All types of anemia	Error rate, Time Taken	Random Forest Tree - 97.57%
DT, Rule Induction, and NB [5]	Hospital data	3 types of anemia	Precision, recall, F-measure	Rule induction- 57%-67% (Predicting Hematology)
Five ensemble methods [6]	500 instances of anemia types (each of 100 types)	Classifying individual RBCs	Min. Value, Max. Value	K-NN and Decision Tree-highest accuracy
K-NN, DT, RF, ANN, SVM [7]	Data from 186 patients	All types of anemia	Age, RBC, MCV, HB, RDW	MCC and AUC of 95.59, 0.87 and 0.98,
C4.5 DT, SVM [8]	514 patients	Classifying individual	F-measure,	C4.5 decision tree- 99.42 %
BN, NB, LR, MLP [9]	539 patients	All types of anemia	Info. Gain, Correlation, OneR, Symmetrical Uncert	LR- 87.3% MPL- 87.1%
K-means [10]	1500 patient records	Blood-related diseases	Clusters	
Standard descriptive statistical method [11]	1,356 Records	Iron deficiency	BTT screening	35 subjects with HbA2 between 3.2 and 3.5% were at risk of false (-) diagnosis of BTT
Causal discovery algorithm [12]	IDA dataset	Iron deficiency	Clinical parameters	Two kinds of relations (("X → Y") and ("X ↔Y"))
SVM [13]	Red blood cells abnormalities	Different kinds of anemia	Image identified	Accuracy of 93.33%
K-NN, C 4.5, ANN[14]	50-Healthy, 50-patients	Iron deficiency anemia	Biochemical parameters	ANN- 78.31
DT, Association Rule mining [15]	Dataset- NFHS-4	Blood-related Anemia	Support, Confidence, Lift	DT -97.35%
DT, SVM and KNN [16]	104 participants	Anemic condition and other diseases	By using image	Accuracy of 82.61%
ANNs, SVMs, NB and ensemble decision tree [17]	IDA Dataset	Different kind of anemia	Information gain, information gain ratio, PCA and CFS	Bagged decision tree-85.6%

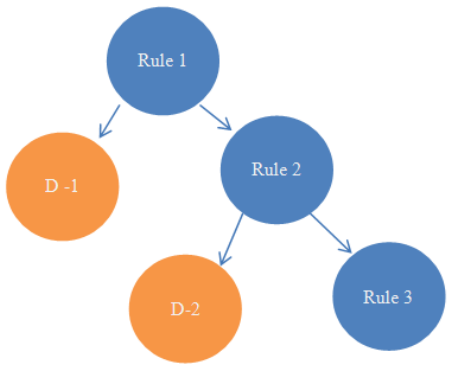

Figure 7: Decision Tree

Use of a library for web scraping

- **Selenium:** is a freely-accessible library for automatic evaluation. It is employed to monitor portal activity. Enter the following line in your terminal to install this library.
- **Pandas:** The Pandas library is utilised for data analysis and manipulation. The data is extracted and stored using it in the desired format.
- **Beautiful Soup:** Utilised to extract data from HTML and XML files, BeautifulSoup is a Python library. It's intended mostly for web scraping. It collaborates with the parser to offer a natural way to browse, search, and alter the parse tree.
- Using BeautifulSoup, implement web scraping in python

Part 1: loading web pages using the request method
Part 2: Title extraction using BeautifulSoup

Part 3: Soup-ed body and head
Part 4: Select with BeautifulSoup
Part 5:Items being scraped right now
Part 6: Extracting Links
Part 7: Generating CSV from data
For classification

Decision Tree (DT)

It is a machine learning application-specific without a parametric supervised learning model with a tree-like topology. The system is built to work from the top down and each internal node has an input feature indicated on it. Each leaf node has a label that identifies the class that was used to forecast the desired variable. Tree splitting is necessary for the DT to keep the prediction class.

Support Vector Machines (SVM)

A type of machine learning technique called Support Vector Machines (SVM) makes use of support vectors to learn new data. A supervised machine learning method called the Support Vector Machine (SVM) can be applied to both classification and regression tasks. A supervised machine learning algorithm is another name for it. Following data classification, we identify the hyper-plane that most effectively distinguishes the two groups.

Naive Bayes (NB)

The likelihood for each class occurring is determined by the probabilistic method known as NB. The classifier facilitates finding the probability of anticipated classifications.

Artificial neural networks (ANNs)

A computational network based on biological neural networks, which create the human brain's structure, is typically referred to as an artificial neural network. Artificial neural networks also feature neurons that are linked to one other in different layers of the networks, just as neurons in a real brain.

Research Challenges

- The internet is too big. The web is incredibly large and expanding quickly. It appears that the web is too big for data mining and data warehousing.
- Complexity of online pages: There is no overarching structure to the web pages. When compared to a typical text document, they are exceedingly sophisticated. The online digital library contains a sizable number of documents. There is no specific sort order used to group these libraries.
- The information on the web is a dynamic source because it is frequently updated. The information is updated frequently, including the news, stock markets, weather, sports, and shopping.
- Diverse user communities: The online user community is growing quickly. The backgrounds, interests, and usage objectives of these users vary. More over 100 million workstations are currently online, and that number is growing quickly.

Relevancy of Information: It is believed that an individual is typically only interested in a small section of the internet, while the remaining amount of the internet provides information that is irrelevant to the user and may obstruct desired results.

Conclusion

Data mining techniques are used in this work to extract expert knowledge on the healthcare sector, in particular, the blood-related disease anemia. Because there is a lot of work to be done and this is a huge subject, we thought that this study would be useful for highlighting areas that could benefit from future research. The suggested method is based on pattern recognition, which is the cornerstone of many sophisticated diagnosis systems. Decision trees, which are important in data mining, have been used in the feature extraction and classification phases of the pattern recognition process there. As a result, a decision support system is projected using pattern recognition, and its functionality is ascertained using data mining approaches. The developed decision-support system will be very helpful for decision-making and the healthcare organization.

Reference

[1] Dogan, Sengul&Turkoglu, Ibrahim. (2008). Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees. International Journal of Science & Technology Volume. 3. 85-92.

[2] S. J. Mohammed MOHAMMED, A. A. Ahmed, A. A. Ahmad and M. Sami MOHAMMED, "Anemia Prediction Based on Rule Classification," 2020 13th International Conference on Developments in eSystemsEngineering (DeSE), 2020, pp. 427-431, doi: 10.1109/DeSE51703.2020.9450234.

[3] Abdullah, Manal& Al-Asmari, Salma. (2016). Anemia types prediction based on data mining classification algorithms.

[4] Sasikala, N. &Banu, Gulmohamed&Babiker, Thgani&Rajpoot, Pushp. (2021). A Role of Data Mining Techniques to Predict Anemia Disease. International Journal of Computer Applications. 174. 16-20. 10.5120/ijca2021921090.

[5] A. H. Shurrab and A. Y. A. Maghari, "Blood diseases detection using data mining techniques," 2017 8th International Conference on Information Technology (ICIT), 2017, pp. 625-631, doi: 10.1109/ICITECH.2017.8079917.

[6] P. T. Dalvi and N. Vernekar, "Anemia detection using ensemble learning techniques and statistical models," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2016, pp. 1747-1751, doi: 10.1109/RTEICT.2016.7808133.

[7] Laengsri, V., Shoombuatong, W., Adirojananon, W. *et al.* ThalPred: a web-based prediction tool for discriminating thalassemia trait and iron deficiency anemia. *BMC Med Inform DecisMak* **19**, 212 (2019).

[8] Sanap, S.A., Nagori, M., Kshirsagar, V. (2011). Classification of Anemia Using Data Mining Techniques. In: Panigrahi, B.K., Suganthan, P.N., Das, S., Satapathy, S.C. (eds) Swarm, Evolutionary, and Memetic Computing. SEMCCO 2011.

Lecture Notes in Computer Science, vol 7077. Springer, Berlin, Heidelberg.

[9] M. S. MOHAMMED, A. A. AHMAD and M. SARI, "Analysis of Anemia Using Data Mining Techniques with Risk Factors Specification," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154063.

[10] Dinakaran, K. &Preethi, R.. (2013). A Novel Approach to Uncover the Patient Blood Related Diseases using Data Mining Techniques. Journal of Medical Sciences(Faisalabad). 13. 95-102. 10.3923/jms.2013.95.102.

[11] SrdjanDenic, Mukesh M. Agarwal, Bayan Al Dabbagh, Awad El Essa, Mohamed Takala, SaadShowqi, JavedYassin, "Hemoglobin A_2 Lowered by Iron Deficiency and α-Thalassemia: Should Screening Recommendation for β-Thalassemia Change?", *International Scholarly Research Notices*, vol. 2013, Article ID 858294, 5 pages, 2013.

[12] F. A. Nugroho, T. H. A. Ederveen, A. Wibowo, J. Boekhorst, M. I. de Jonge and T. Heskes, "Application of A Causal Discovery Model to Study The Effect of Iron Supplementation in Children with Iron Deficiency Anemia," 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), 2019, pp. 1-5, doi: 10.1109/ICICoS48119.2019.8982503.

[13] C. C. Hortinela, J. R. Balbin, J. C. Fausto, P. Daniel C. Divina and J. P. T. Felices, "Identification of Abnormal Red Blood Cells and Diagnosing Specific Types of Anemia Using Image Processing and Support Vector Machine," 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2019, pp. 1-6, doi: 10.1109/HNICEM48295.2019.9072904.

[14] .T. İlaslaner and A. Güven, "Investigation of the Effects Biochemistry on Iron Deficiency Anemia," 2019 Medical Technologies Congress (TIPTEKNO), 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895227.

[15] .Meena K, Tayal DK, Gupta V, Fatima A. Using classification techniques for statistical analysis of anemia. ArtifIntell Med. 2019 Mar;94:138-152. doi: 10.1016/j.artmed.2019.02.005. Epub 2019 Feb 19. PMID: 30871679.

[16] . Noor, Nahiyan& Anwar, Mohammad &Dey, Mrinmoy. (2019). Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition. 10.1109/BECITHCON48839.2019.9063188.

[17] K.Y., Tuba &Yurtay, Nilüfer&Öneç, Birgül. (2021). Classifying anemia types using artificial learning methods. Engineering Science and Technology, an International Journal. 24. 50-70. 10.1016/j.jestch.2020.12.003.