NLP And Conformal Prediction Reliably Filter Drug-Induced Liver Injury

[1] SAU KRISHNA PIPPALLA, [2] CH. SURESH

[1] MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280

saikrishna.pippalla@gmail.com

[2,] Assistant Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram, Narsapur, Andhra Pradesh 534280

***Abstract**: Drug-brought on liver injury describes the side outcomes of drugs that damage the liver. Life-threatening outcomes have additionally been said in severe instances. Liver toxicity is consequently a critical take a look at for brand new drug users. These reports are documented inside the medical literature containing preliminary in vitro and in vivo experiments. Traditionally, extracting information from publications is based on using textual information, which limits the overall performance of information extraction. The improvement of natural language processing permits for the automated processing of biomedical documents. Here, primarily based on about 28,000 articles (titles and abstracts) furnished with the aid of the Critical Review of Big Data Research, this studies is in comparison to the possible models. Right at filtering hard records. Among the five techniques of writing the textual content, the model using frequency-converted data (TF-IDF) and logistic regression accomplished higher than others with an accuracy of zero.957 of the validation manner. In addition, a combined version with ordinary efficiency turned into created with a logistic regression version of the predicted possibility derived from specific models with unique factorization methods. The cluster model completed an accuracy of 0.954 and an F1 score of zero.955 within the reference records saved as part of the undertaking. In addition, vast fantastic/bad predictors had been diagnosed by way of sample interpretation. Expectations of forecasting are covered in forecast compliance, which lets in users to control forecast uncertainty. Overall, the proposed model and the TF-IDF model completed proper classification results, which can be utilized by researchers to fast clear out statistics that describe the situations within the heart is harm.*

Keywords: Drug-induced liver injury, natural language processing, ensemble learning, sentence embedding,conformal prediction.

I. INTRODUCTION

DURG-precipitated liver harm (DILI) is defined as a negative reaction of the liver to the drug. DILI is a not unusual and essential motive of liver damage because the liver plays a critical position in drug metabolism. Liver toxicity due to tablets may be divided into two kinds: intrinsic and idiosyncratic. High intrinsic drug toxicity is greater predictable and is directly related to the dosage of specific capsules. Liver damage takes place within a short period of time, usually within some hours of taking the medicine. In comparison, idiosyncratic liver toxicity is more patient precise and looks longer. For capsules with excessive lipophilicity, idiosyncratic liver harm may additionally occur even under the endorsed daily intake. The severity of DILI can range extensively

Sufferers bear in mind the interplay among genetic and environmental factors. Although maximum sufferers can get over DILI, acute DILI can lead to liver failure. For example, paracetamol liver toxicity, often as a result of overdose, is concept to be responsible for seventy three.7% of liver disorder and liver failure in Scotland among 1992 and 2014. Additionally, around 75% of idiosyncratic reactions bring about transplantation hepatic or loss of life. Therefore, DILI has end up one of the maximum common motives for rejection of latest drug applicants and is fastidiously evaluated at some stage in drug development. The complex mechanisms of DILI and the seriousness of its effects require higher tracking of DILI instances. However, most people of DILI reviews come from medical trials or medical trials posted within the free literature. By law, scientific courses should be checked and processed by way of scientists and pharmacists.

II LITERATURE REVIEW

1. "Drug-induced liver damage," Nature Rev. Dis. Primers"

AUTHOR:" R.J. Andrade"

Drug-induced liver damage (DILI) is an adverse reaction that occurs when you take capsules or other xenobiotics. It can occur as a predictable occurrence as a man or woman gets exposed to harmful amounts of some chemicals or in an unpredictability event with a variety of tablets used in daily use. The effects of drugs may be detrimental for the liver of individuals due to genetic or environmental risk factors. The risk factors affect hepatic metabolism and the elimination of the DILI-causative agents, causing cell stress, loss of vitality, activation of an adaptive immune reaction as well as a lack of development in a way that leads to complete damage to the liver. The idiosyncratic DILI is an extremely rare illness of the liver, however it could be extremely severe and, sometimes, fatal. It is characterized by several phenotypes that mimic various hepatic illnesses. The recognition of DILI is determined by being able to exclude other causes of liver disease, as specific biomarkers remain unidentified. The clinical scales that comprise CIOMS/RUCAM could aid in the identification procedure but require improvement. The variety of clinical parameters that are studied in prospective cohorts could be utilized to determine the severity of DILI result. While no treatment pharmacological is thoroughly examined in controlled scientific studies Corticosteroids are a good option especially in the emerging design of DILI caused by immune-checkpoint inhibitors for cancer patients.

2. Liver Tox"

AUTHOR: J. H. Hoofnagle, J. Serrano, J. E. Knoben, and V. J. Navarro.

The liver damage caused by drugs is among the most difficult forms of liver disease with regards to prognosis as well as treatment. A number of hundred pills, nutritional supplements and medicinal herbs have been linked to causing injuries to the liver. The symptoms they manifest in may be extremely varied and appear to mimic almost any type of liver disease. The amount of research that has been done on the subject of drugs causing liver damage is massive and spread across a variety of journals covering a wide range of disciplines and different languages. The best textbooks can be obtained, but they're outdated and not always accessible. In the case of liver injuries caused by drugs, it's an extremely difficult area of research due to the fact that many instances are erratic, unique and rare, making them consequently hard to study. Because of this that, there was a lack of advancements regarding the technology of manipulation or prevention of the effects of drugs caused liver damage in the last fifty years.

3)"Distributed expressions, words and their composition."

AUTHOR: T. Mikolov, I. Sutskever, K. Chen, G. S. Corridor, and J. Dean.

The recently released Non-stop Skip gram is a reliable method of learning about top-quality assigned vector representations which capture much unique relationship between semantic and syntactic. This paper provides several extensions to improve the performance of the vectors as well as the speed of education. Sub sampling common words, we can achieve huge speedups and also research more regular representations of words. Also, we describe an easy way to use the hierarchical Soft max also known as bad sampling. A fundamental problem with phrases is their insensitivity to order in a phrase and also their ability to represent phrases that are idiomatic. In this case “Canada" and "Air" have different meanings "Canada" and "Air" cannot be easily joined to produce "Air Canada". In the light of this scenario, we provide the simplest method for finding words in textual material, and show that the ability precise vector representations for the tens of millions of words are feasible.

III System Analysis

FEASIBILITY STUDY

The possibility of the project will be assessed in this stage and the enterprise's vision is set ahead with a totally modern design for the venture along with a couple of cost estimations. When evaluating a machine, the feasibility check of the machine proposed planned to be carried out. This will ensure that the proposed device will not be a burden for the company. To conduct feasibility study and knowledge of the main requirements to the equipment is crucial.

Three of the most important issues in the feasibility assessment three of the most important issues in feasibility analysis

* ECONOMICAL FEASIBILITY

* TECHNICAL FEASIBILITY

* SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

The purpose of this look-up is by evaluating the economic impact that the machine could have upon the company. The amount that an organization can invest in the development and research of the device is restricted. The costs must be able to justify the costs. This is why the latest gadgets are also within the cost budget and was completed since the vast majority of the tech used is readily available. The only thing that was custom-designed required to be bought.

TECHNICAL FEASIBILITY

The examination will be conducted to determine the feasibility of technology, that is, the technical specifications for the device. The gadget that is developed must be able to meet the minimum demand on available technology sources. This can lead to increased demand on available technological sources. It will result in the putting of a lot of demands upon the client. The new system should have only a small amount of requirement as the smallest of adjustments, or none at all can be made to the device.

SOCIAL FEASIBILITY

The goal of research is to determine the degree of a system's popularity by involving the users. It is also a process of educating the user on how on how to utilize the device effectively. Users should not be intimidated by the device. As the alternative is to treat the device as a necessity. The level of acceptance the assistance of the user is contingent upon the strategies that are employed in order to teach the user about the gadget and get them familiar with the device. Its self-confidence is to be built up so that he's competent to offer a couple of constructive critiques, which are welcomed because he's ultimately the customer of the gadget.

IV Data Set Description

1. Title:

It is able to filter drug-induced liver injuries Literature using Natural Language Processing and conformal Prediction

2. Description:

The data you've mentioned is likely to have been a result of screening out liver damage from drugs research by using Natural Language Processing (NLP) and methods of conformal prediction. Based on the name the data suggests that the data set could contain the data on liver injury caused by drug (DILI) and may be composed of text-based records that include medical research, medical information or database of drugs. Natural Language Processing is a field of artificial intelligence aimed on the interplay between computers and human languages. It involves analyzing and processing vast quantities of information about herbal languages together with textual data in order to gain meaningful insight or carry out precise tasks.

Conformal prediction, on contrary is a device for learning method which provides measures of the level of confidence in the predictions made with the version. This is particularly useful in situations where the credibility of predictions is essential and is also essential in scientific or medical applications. Based upon the description of the dataset it appears like the aim is to create an approach that makes use that makes use of NLP and conformal predictions to effectively filter the literature pertaining to the liver-harm caused by a drug. This could be beneficial for medical professionals, researchers as well as pharmaceutical companies to be informed of pertinent records and take informed choices regarding medication safety.

From the information provided in the head notes Based on the side headings, it appears that you're looking for an explanation or definition of the data description, which is "Reliably Filter Drug-Induced Liver Injury Literature with Natural Language Processing and Conformal Prediction." This is a possible description of the various phases that might entail:

"Introduction": This portion typically provides a high-level description of the issues and the motivation for conducting an examination and the importance of ensuring that drugs are filtered out from research on liver damage.

Background: The authors could talk about of the current techniques or challenging circumstances in the filtering of literature that are associated with the liver injuries caused by drug precipitation. It should also include guidance methods for creation, the limitations of current methods and the advantages that can be derived from organic language processing techniques and methods for conformal prediction.

Methodology: This section could be able to dive into the details regarding how herbal processing and conformal predictions are employed to sort the research literature. This could include sections that include:

* Pre-processing and data series explain where the data comes from as well as the method of organizing it for study.

Natural language processing methods description of NLP techniques employed. This includes tokenization of textual content as well as popularity of the named entity or sentiment assessment.

Framework for conformal prediction: A description of how conformal prediction processes are implemented to determine the validity of predictions in the basis of this situation.

Design

Method for Reliable Filtering the literature on Liver Injuries caused by drugs using natural Language Processing and a Conformal Prediction

1. Introduction

Damage to the liver caused by drugs (DILI) is a major problem in pharmacy co vigilance and requires efficient ways to determine relevant studies. This paper suggests a technique that combines NLP (NLP) as well as conformal prediction to effectively filter out DILI-related research.

2. Methodology

Use conformal prediction for uncertainty estimation: Employ the conformal prediction technique, which is a helps you understand the framework to estimate uncertainty when categorizing documents as DILI-related. Provide valid confidence indicators for the predictions in order to increase their reliability.

3. Evaluation

Measure performance by using of precision. Also take consideration, and assess the accuracy of the model of conformal prediction.

Comparative Analysis: Compare the current method to show superior efficiency and accuracy.

4. Results

* **Performance Evaluation**: Higher precision and recall compared to baseline techniques imply effectiveness. Well-calibrated uncertainty estimates decorate trustworthiness.

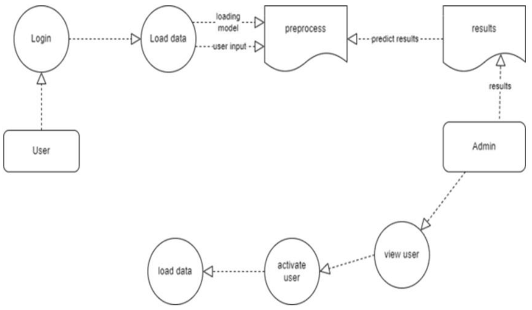
5. Conclusion

Summary: Presented a unique method of reliable filtering the DILI literature.

Resolved the challenges of traditional methods with the help of increasing the accuracy of techniques and confidence in self.

Future Directions: Extend the method to different adverse drug reactions.

Add additional capabilities to improve performance. This design outlines a comprehensive method for ensuring that DILI is properly filtered from research, assisting in the proactive efforts of pharmacy co vigilance.



V MACHINE LEARNING ALGORITHMS

Decision tree classifiers

Classifiers of decision tree are employed effectively in a variety of sectors. They are primarily used for their ability to capture photos of descriptive information about decision-making in the form of information. The decision tree is generated through educational units. The method for this technology is based on the complete number of items (S) each belonging to a specific learning C1 and C2 ..., Ck, is in the following order:

Step 1. If all the devices included in S are in the same class, like Ci the choice tree of S includes a leaf labeled in this manner.

Step 2. If not, then let T be an examination that has a possibility of achieving outcome O1 and O2,,..., O2, and so on. Each of the items within S will have one final outcome for T. Therefore, the check is a wall that divides S into subsets: S1 and S2. ,... Sn, where each object within Si is the final result Oi in T. T is the basis of the tree selection and for each result Oi we create a secondary decision tree through the exact procedure again in the Set Si.

Gradient boosting

Gradient boosting is a machine-learning method used in category and regression work, as well as other tasks. It is a variant of prediction which is in the form of an assortment of precarious methods of prediction, which could be typically decision trees. When a tree of choice is the weakest learner that results, it is referred to as gradient-boosted timber and typically performs better than random wooded areas. The gradient-boosted timber model is constructed using a level-clever method similar to the various boosting methods however; it extends the alternative techniques to allow the optimization of any differentiable loss feature.

K-Nearest Neighbors (KNN)

Simple, but a completely effective class algorithm Classifies based totally on a similarity measure Non-parametric Lazy learning Does no longer "study" until the take a look at instance is given When ever we've a new statistics to categories, we discover its K-nearest neighbors from the schooling records

Example

Training dataset has good-quality instances of the feature area method, space for categorization variables (non-metric variables)Learning mostly based upon experiences, and as a reason it also performs poorly because an event within the vector of entry for taking a look or prediction could be delayed in the data set for schools.

Logistic Regression Classifiers

Logistic regression analysis studies the connection between categorical variables along with a quick and easy analysis of the independent (explanatory) variable. Logistic regression is used in cases where the structured variable is found to have most desirable values, in addition to the two variables 0 and 1, or Yes and Yes and. The term multinomial logistic regression is usually reserved for cases where the dependent variable contains more than three unique or distinctive values for example, divorced, single, married or Widowed. Even though the kind of data that are used to determine the base on variable may differ from the case of multiple regressions, its use in the real world procedure is comparable. Logistic regression is akin to discriminate analysis in the context of a method to analyze expression-reaction variables. A lot of statisticians believe how logistic regression can be more adaptable and better suited to the analysis of most scenarios than discriminate evaluation. This is due to the fact that it doesn't rely on the fact the fact that independent variables are evenly distributed as discriminate evaluation does.

The software calculates bi-logistic regression as well as multinomial logistic regression using the numeric as well as categorical independent variables. It analyzes the equation of regression as and the accuracy of the odds ratio, match of confidence, probability limits as well as deviance. It provides a comprehensive review of the residuals such as review of diagnostic residuals and graphs. It is able to perform an independent subset selection search to find the best regression model using most impartial variables. The model provides self-assurance intervals for anticipated results and provides ROC curves to help you to determine the best cut off to determine the quality cut off factor for each category. This allows you to confirm your findings by automated classification of rows not being used during the course of the test.

Naive Bayes:

The Bayes naive approach is a monitored approach to getting to know that is founded on an unproven theory: that the absence (or absence) of a particular function of a specific category does not correspond to the existence (or the absence) of any other characteristic.

But, despite the case, it appears to be robust and green. The performance of the device is similar with other supervised mastering methods. Different motives are superior in the research literature. In this article, we focus on a theory based solely on the bias of representation. The naive Bayes classifier is a classifier that's linear and also a linear discriminate model, also known as logistic regression, or an SVM that is linear (aid vector device). The distinction lies in how you estimate what the parameter of the classification (the learn bias).

Although the Naive Bayes classifier has been widely utilized in the research world however, it's not always fully-sized for those who need to get results that are usable. One of the reasons is that the scientists have discovered that it is extremely easy to set up and practice, the parameters are simple to determine and mastering can be extremely quick even with huge databases. Its accuracy is and fairness is a good assessment to other procedures. However most customers are not able to create an understanding of the model that is easy to understand and implement, as they are no longer able to appreciate the fascination of such a method.

We present an updated presentation of what we have learned in this way. It is easier to comprehend and the way it is used also becomes less complex. In the beginning of this tutorial review, we discuss some theoretic aspects of the naive Bayes classifier. After that, we apply the method on a set of data that is analyzed using Tanagra. Then, we compare the results (the variables of the model) with respect to the results obtained using different linear methods that include the logistic regression, linear discriminate analysis, and the linear SVM. It is important to note that the effects are extremely steady. This is largely the reason for the excellent overall performance of this technique for evaluating other. In the second portion, we utilize a range of methods with the same data (Weka 3.6.0), R 2.9.2 2.1.1, Knime 2.1.1 3.1.1, and 2.0band Rapid Miner 4.6.Zero, 2.0band Rapid Miner 4.6.Zero). Our goal is to specifically understand the resulting effects.

Random Forest

Random forests, also known as random choice forests is an ensemble learning to be aware method that is used for classifying Regression, classification and various tasks that works by creating the appearance of decision bushes during the time of training. When it comes to classification the result of the random forest is the classification that is selected using the most forests. Regression responsibilities require that the median or average estimation of the individual timber is returned. Random choice forests are accurate in addressing the decision-making bushes' habit of being over fitted to their schooling sets. Random forests typically outperform the choice timber However; their accuracy is less than gradient-boosted trees. But, the characteristics of statistics could affect their performance.

SVM

When performing category tasks, the discriminate method of mastering seeks to identify the basis for an unrelated and equally allotted (iid) training data set the discriminate characteristic that will correctly anticipate labels for newly obtained instances. Contrary to the generative system-learning processes that require the computation of conditional probability distributions, the discriminate class algorithm takes the records' number x as a factor and applies it and various instructions that are a part of the category-related task. Less powerful than generative approaches, which can be generally used when prediction entails outlier detection, discriminate approaches require fewer computational assets and much less schooling information, specifically for a multidimensional feature space and whilst only posterior chances are wanted. From a geometric perspective the process of learning a classifier is similar to finding the equation that defines an equilateral floor with a quality that separates outstanding classes within the functional area.

The SVM technique is known as a discriminate approach which, because it resolves the convex optimizing issue analytically, it continuously provides the exact gold-standard hyper plane parameters--as opposed the genetic algorithm (GAs) as well as perceptions each of which is extensively used in category research on gadgets. The solutions for perceptions specifically depend on parameters for the initialization and ending. A specific kernel transforms the data from the input space into the typical area, learning provides a unique set of SVM model parameters that are specific to an educational set as perception and GA classification models are distinctive for each occasion that when training begins. The main purpose of GAs as well as perceptions is best in reducing mistakes that occur at any moment in education and to convert into a number of hyper planes' that satisfy this need.

VI MACHINE LEARNING ALGORITHMS

To effectively filter out harmful liver effects caused by drugs (DILI) research, using an appropriate device and knowing the latest trends is essential. An entire approach includes a variety of methods to guarantee the efficacy of these methods. Initial, pre-processing procedures for records are carried out to simplify and uniformize biomedical text information, ensuring consistency and eliminating the noise. It includes tokenization; forestall word elimination and stemming in order to improve the quality input capabilities. The feature engineering process plays an essential role in obtaining relevant information from texts as well as determining chemical names, harmful effects of liver injuries, and their indications. The use of domain-specific knowledge is frequently used for character extraction and characteristic selection to aid in the creation of useful features to the kind project.

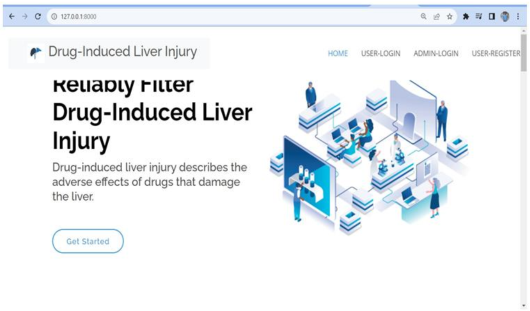
Additionally, mixing NLP and natural process of language (NLP) techniques enhances the ability of the model to find relevant records from textual data. Strategies like recognized entity names (NER) and the tagging of spoken words help in the process of feature extraction and contribute to the model's general precision. It is used to assess uncertainty in the method of classification, which ensures that outputs from models are provided with reliable confidence measurements. This process enhances the credibility of the predictions, and provides insight on the model's self-assurance when it comes to its decisions.

Evaluation methods play an essential function in assessing accuracy and efficiency of machines using the study methods. Cross-validation can be utilized to test models' generalization based on unknown facts providing insights about its reliability. The evaluation of the confusion matrix allows thorough analysis of how the model performs when it comes to determining DILI literature including metrics that include accuracy, recollect and the F1-score of every quality. Calibration plots help determine the validity of predicted scenarios, and ensure that the estimates of uncertainty in the version are accurate.

The final step is the comparison of overall performance metrics for the version such as precision, recall, F1-score and the accuracy of a version, by comparing them with benchmark strategies as well as previous studies. The efficacy of uncertainty estimation using conformal prediction is also assessed, revealing insight into the reliability of the version. With these strategies, model learning can effectively and efficiently clear the drug-caused liver-related harms literature.

RESULTS

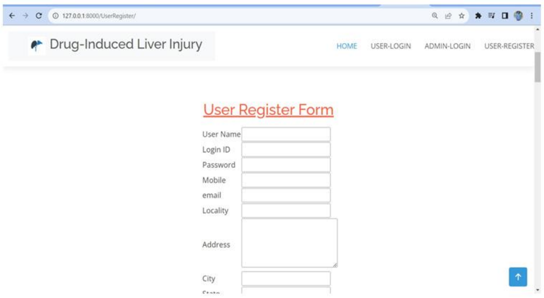
Home Page



Reliably Filter Drug-Induced Liver Injury

Drug-induced liver injury describes the adverse effects of drugs that damage the liver.

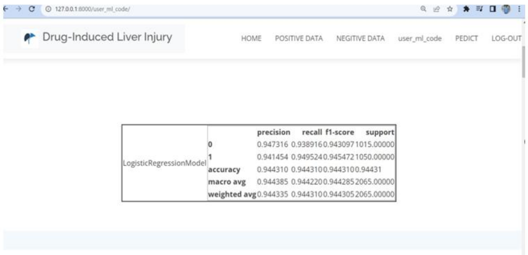
User Register



User Register Form

User Name
Login ID
Password
Mobile
email
Locality
Address
City
State

MI Results



		precision	recall	f1-score	support
LogisticRegressionModel	0	0.947316	0.938916	0.943097	1015.00000
	1	0.941454	0.949524	0.945472	1050.00000
	accuracy	0.944310	0.944310	0.944310	0.94431
	macro avg	0.944385	0.944220	0.944285	2065.00000
	weighted avg	0.944335	0.944310	0.944305	2065.00000

VI CONCLUSION

Different methods were developed to reduce DILI-related data that are based entirely on four strategies for victimization (bag-of-words TF-IDF Word2vec biomedical models, and send2vec biomedical model) as well as clustering. The model that uses TF-IDF and LR was more efficient than the other models having an AUROC being zero.990 and accuracy of 0.957 as well as an AUPRC being zero.990. A model for getting familiar which has the best overall performance, but with an extremely limited range of words is constructed using the ability to predict the probability of twelve human word victimization algorithms; this indicates that it is the most precise (0.954) as well as the F1 score of the data analysis (zero.955). . Both styles perform very effectively on both the records that show every model's main on different metrics (mastering model precision: zero.954, F1 score: zero.955, accuracy: zero.960 don't forget to mention: zero.950). Furthermore, the comply with-up estimations are used to build confidence from the straight outcomes, thus serving as a benchmarking instrument for researchers who want to stay clear of FN estimations . The creation of TF-IDF as well as the mixed model allows users to utilize both these models to reference methods of learning based upon the methods that permit more effective filtering of data without causing any harm to researchers with knowledge of DILI in the field of drug.

REFERENCES

1. M. Chen, J. Borlak, and W. Tong, “High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury,” *Herpetology*, Baltimore, MD, USA, vol. 58, no. 1, pp. 388–396, Jul. 2013.

2. R. J. Andrade et al., “Drug-induced liver injury,” *Nature Rev. Dis. Primers*, vol. 5, no. 1, pp. 1–22, Aug. 2019. [Online]. Available: https://www.nature. Com/articles/s41572-019-0105-0

3. N. Kaplowitz, “Drug-induced liver injury,” *Clin. Infect. Dis.*, vol. 38, no. Supplement_2, pp. S44–S48, Mar. 2004. [Online]. Available: https: //doi.org/10.1086/381446

4. Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

5. M. C. Donnelly, J. S. Davidson, K. Martin, A. Baird, P. C. Hayes, and K. J. Simpson, “Acute liver failure in Scotland: Changes in aetiology and outcomes over time (the Scottish look-back study),” *Alimentary Parma- col. Therapeutics*, vol. 45, no. 6, pp. 833–843, Mar. 2017.

6. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

7. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2019. [Online]. Available: http://arxiv.org/abs/1810.04805

8. Y. Wang, M. Rastegar-Mojarad, R. Komandur-Elayavilli, S. Liu, and H. Liu, “An ensemble model of clinical information extraction and in- formation retrieval for clinical decision support,” in *Proc. TREC Conf. Nat. Inst. Standards Technol.*, 2016, pp. 1–10.

9. Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp: 23-37.

10. X. Zhan, M. Humbert-Droz, P. Mukherjee, and O. Gevaert, “Structuring clinical text with AI: Old versus new natural language processing techniques evaluated on eight common cardiovascular diseases,” *Patterns*, vol. 2, no. 7, Jul. 2021, Art. No. 100289. [Online]. Available: https://www.cell.com/patterns/abstract/S2666-3899(21)001227.

11. Prasadu Peddi (2018), “A STUDY FOR BIG DATA USING DISSEMINATED FUZZY DECISION TREES”, ISSN: 2366- 1313, Vol 3, issue 2, pp:46-57.