An Analysis of blood donors and Hepatitis C patients by using big data techniques

M. Sivachandran [1], Dr.T.Krishnakumar [2]
{sivachandranm08@gmail.com[1], drkk@bharathuniv.ac.in[2]}

Research Scholar, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India.[1],
Professor, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India.[2]

Abstract. The system focuses and produces the optimal solution. Such as the SGD Text approach gives the highest and optimal result such as 88.38% compare with other models. and Simple Linear Regression approach gives very lowest result compare with other models. SGD Text approach produces highest precision value level which is 82.61% compare with other models, The lowest precision value is 60.02% which is produced by Simple Linear Regression approach, the highest precision value is 81.19% which is SGD Text approach. The Simple Linear Regression, SMOreg and SMO and have respectively 0.21,0.23 and 0.25 seconds to build the model. Fast and exact clinical screening is essential for the fruitful treatment of infections. Utilizing AI calculations and dependent on research center blood test results. This information extends the model's utility for use by broad professionals and demonstrates that blood test results contain more data than doctors for the most part perceive.

Keywords: SGDText, SimpleLinearRegression, SimpleLogistic, SMO, SMOreg and VotedPerceptron.

1 Introduction

Evaluation of this danger requires ideal, precise and dynamic blend of the enormous measure of clinical data in the preoperative period. [1-4]Current preoperative danger definition is restricted to a doctor's emotional danger evaluation or danger scores that frequently require expound information extraction [5-10]. While most of existing preoperative AKI hazard scores are restricted to heart medical procedure and have humble precision [11-13], instruments for preoperative danger delineation for extreme sepsis are missing.[14]

Multivariate relapse models are customarily utilized for hazard forecast in clinical exploration because of their simplicity of result understanding and investigation however AI classifiers have picked up energy in biomedical examination during the previous few years with the accessibility of electronic wellbeing records and more perplexing clinical data.[15] Even however the decision of danger expectation model assumes a part in creating vigorous and exact danger prediction,[16] information cleaning and preprocessing are similarly significant for model execution [17].

I3CAC 2021, June 07-08, Chennai, India

DOI 10.4108/eai.7-6-2021.2308658

2 Material And Methods

The dataset collected from UCI repository. The study data set comprise of laboratory diagnosis values of blood donors for the patients and subjects of Hepatitis C patients and more details in demographic values. The below information have given about the list of the attributes.

S.No	**Attribute**
1	X (Patient ID/No.)
2	Category (0=Blood Donor, 0s=suspect Blood Donor, 1=Hepatitis, 2=Fibrosis, 3=Cirrhosis)
3	Age
4	Sex (female,male)
5	ALB
6	ALP
7	ALT
8	AST
9	BIL
10	CHE
11	CHOL
12	CREA
13	GGT
14	PROT

The Weka 3.8.9 has implemented to get the optimal solution of the above dataset. The below approaches have implemented and got optimal solution.

- SGDText,
- SimpleLinearRegression,
- SimpleLogistic,
- SMO,
- SMOreg and VotedPerceptron

3 Results And Discussions

In this section discuss about the results and discussions of this research work. The below table clearly demonstrates that the Accuracy levels of all approaches namely SGDText, SimpleLinearRegression, SimpleLogistic, SMO, SMOreg and VotedPerceptron All of these algorithms belong to Function Category.

Table 1: Various Approaches Vs Accuracy

S.No	Algorithm	Accuracy
1	SGDText	88.38%
2	SimpleLinearRegression	60.02%
3	SimpleLogistic	60.93%
4	SMO	69.23%
5	SMOreg	80.14%
6	VotedPerceptron	77.57%

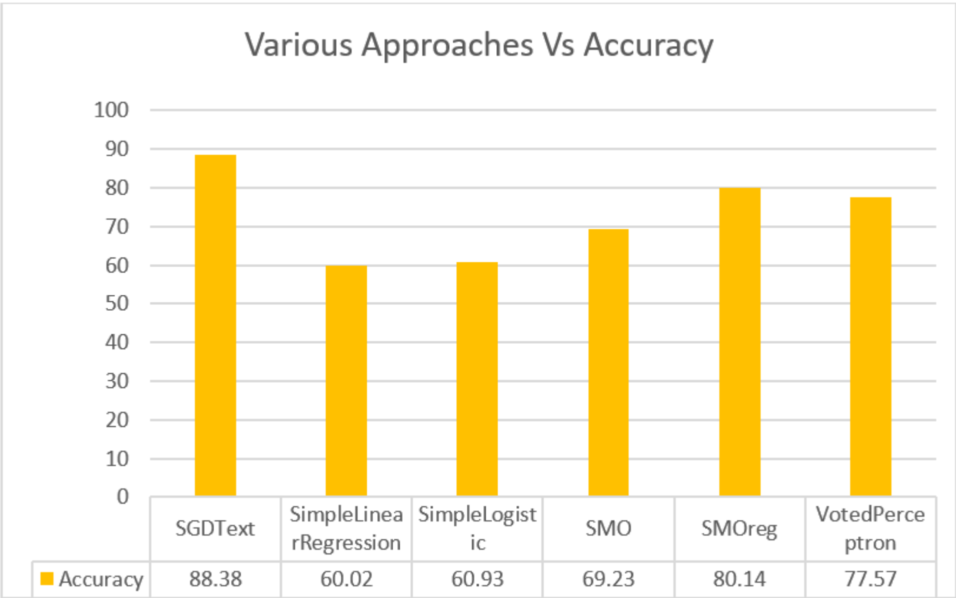

Figure 1: Various Approaches Vs Accuracy

This above diagram clearly represents that the SGDText approach produces 83.38 % of accuracy level , SimpleLinearRegression approach holds 60.02% of accuracy level, SimpleLogistic approach is holding 60.93% of accuracy level, SMO gives the accuracy level is 69.23% of accuracy level, SMOreg approach gives that 80.14% of accuracy level and VotedPerceptron approach has 77.57% of accuracy level.

The SGDText approach and SMOreg approach have above 80% accuracy level, VotedPerceptron has 77.57% and rest of the SimpleLinearRegression approach, SimpleLogistic approach, and SMO approach have the range between 60% to 70% .

The SGDText approach gives the highest and optimal result such as 88.38% compare with other models. and SimpleLinearRegression approach gives very lowest result compare with other models.

Table 2: Various Approaches Vs Precision Values

S.No	Algorithm	Precision
1	SGDText	82.61%
2	SimpleLinearRegression	66.8%
3	SimpleLogistic	67.57%
4	SMO	62.7%
5	SMOreg	81%
6	VotedPerceptron	70.68%

The above table clearly demonstrates that the Precision levels of all approaches namely SGDText, SimpleLinearRegression, SimpleLogistic, SMO, SMOreg and VotedPerceptron All of these algorithms belong to Function Category.

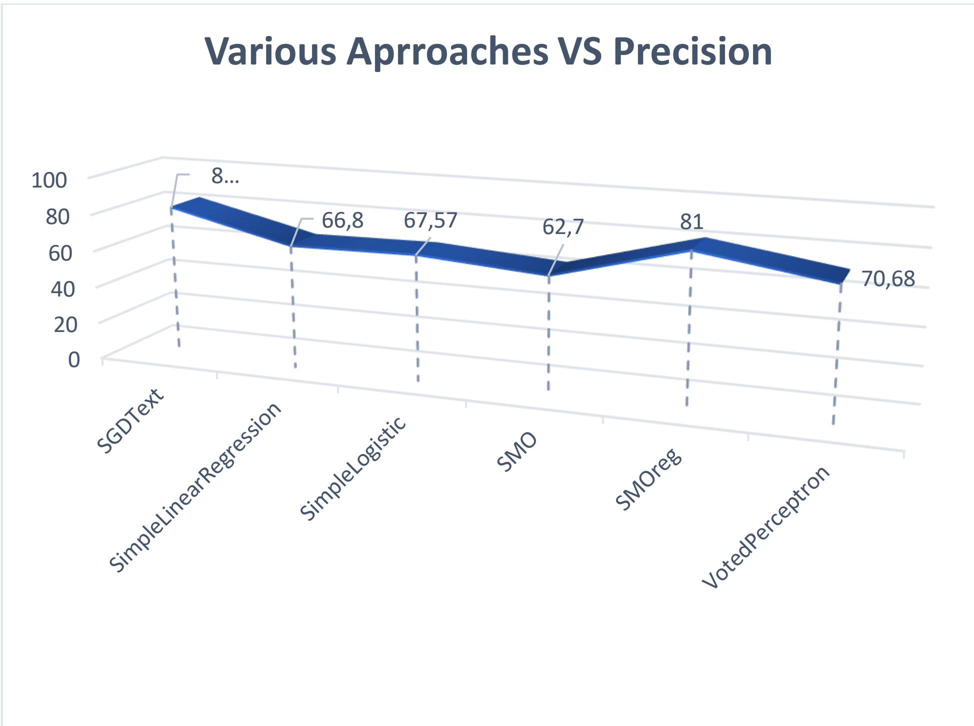

Figure 2: Various Approaches Vs Precision

The above diagram represents that the SGDText approach produces 82.61% of precision value, SimpleLinearRegression approach holds 66.8% precision value, SimpleLogistic approach is holding 67.57% precision value, SMO gives the precision value approach is 62.7% , SMOreg approach gives that 81% precision value and VotedPerceptron approach has 70.68% of precision value.

SGDText approach produces highest precision value level which is 82.61% compare with other models, Next highest precision value is produced by the SMOreg approach which is

81%. The next highest priority is VetedPerception approach which is 70.68% of precision value, the rest of the SimpleLinearRegression, SimpleLogistic and SMO have the range between is 60% to 67%.

Table 3: Various Approaches Vs Recall Values

S.No	Algorithm	Recall
1	SGDText	81.19%
2	SimpleLinearRegression	60.02%
3	SimpleLogistic	60.68%
4	SMO	62.49%
5	SMOreg	80.14%
6	VotedPerceptron	71%

The above table clearly demonstrates that the recall values of all approaches namely SGDText, SimpleLinearRegression, SimpleLogistic, SMO, SMOreg and VotedPerceptron All of these algorithms belong to Function Category.

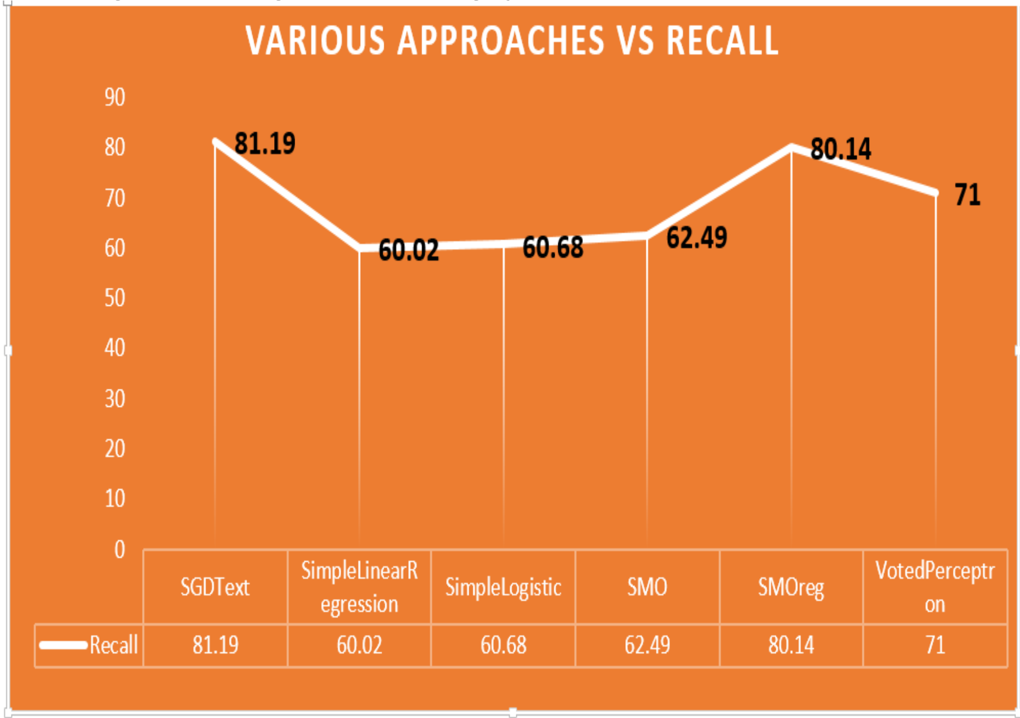

Figure 3: Various Approaches Vs Recall Values

The above diagram represents that recall values have been produced by using various algorithm the SGDText approach produces 81.19% of Recall value, SimpleLinearRegression approach holds 60.02%% of recall value, SimpleLogistic approach is holding 60.68% of recall value, SMO approach gives the recall value is 62.49% , SMOreg approach gives that 80.14% recall value and VotedPerceptron approach has 71% of recall value.

The lowest precision value is 60.02% which is produced by SimpleLinearRegression approach, the highest precision value is 81.19% which is SGDText approach. The SMOreg approach is 80.14% which is the next highest precision level. The SimpleLinearRegression, SimpleLogistic, and SMO have the range between 60% to 63%.

Table 4: Various Approaches Vs Time Taken to build the model

S.No	Algorithm	Time taken to build model(In Seconds)
1	SGDText	0.19
2	SimpleLinearRegression	0.21
3	SimpleLogistic	0.49
4	SMO	0.25
5	SMOreg	0.23
6	VotedPerceptron	0.91

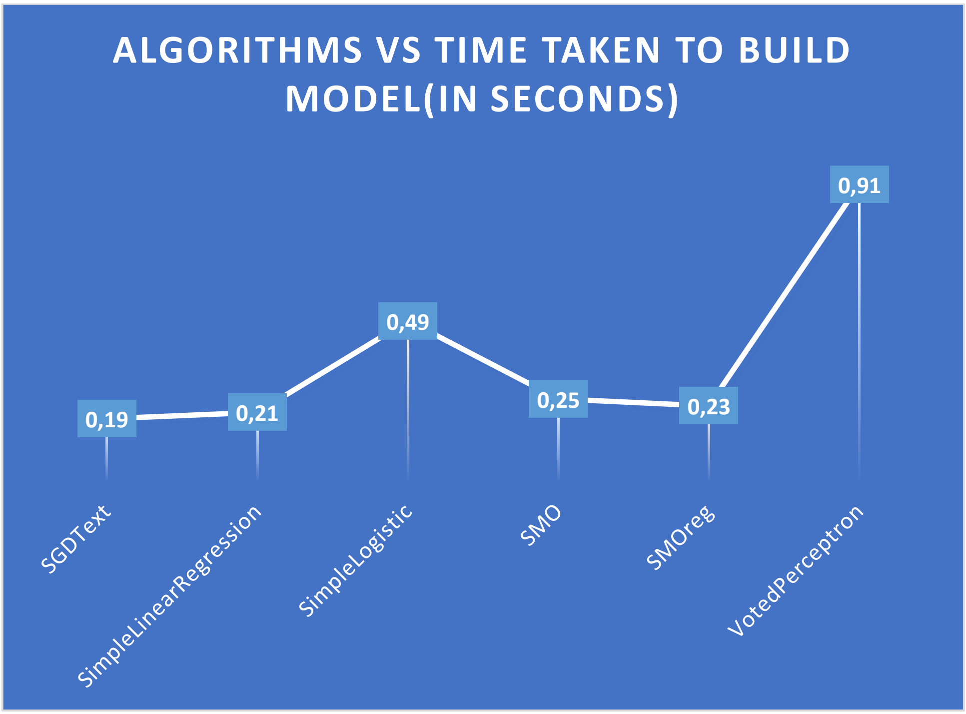

Figure 4: Various Approaches Vs Time Taken to build the models(In Seconds)

The above table clearly demonstrates that the time consumptions of various approaches namely SGDText, SimpleLinearRegression, SimpleLogistic, SMO, SMOreg and VotedPerceptron. All of these algorithms belong to Function Category.

The above diagram represents that all approaches have taken the time to build the model like SGDText approach takes the time to build the model around 0.19 seconds, SimpleLinearRegression approach takes 0.21 seconds to build the model, SimpleLogistic approach takes the time to build the model around 0.49 seconds, SMO approach takes 0.25 seconds to build the model, SMOreg approach takes 0.23 seconds to build the model and VotedPerceptron approach takes 0.91 seconds to build the model.

The SGDText is taking low time consumption to build the model. It takes only 0.19 seconds. It is very low time consumption compare with other approaches for building the models.

The SimpleLinearRegression, SMOreg and SMO and have respectively 0.21,0.23 and 0.25 seconds to build the model. The SimpleLogistic approach has 0.49 seconds to build the model. The highest time has taken VotedPerceptron approach which is 0.91 seconds to build the model.

4 Conclusion

This system concludes that the SGDText approach gives the highest and optimal result such as 88.38% compare with other models. and SimpleLinearRegression approach gives very lowest result compare with other models. SGDText approach produces highest precision value level which is 82.61% compare with other models, The lowest precision value is 60.02% which is produced by SimpleLinearRegression approach, the highest precision value is 81.19% which is SGDText approach. The SimpleLinearRegression, SMOreg and SMO and have respectively 0.21,0.23 and 0.25 seconds to build the model.

References

[1] Khan NA, Quan H, Bugar JM, Lemaire JB, Brant R, Ghali WA. Association of postoperative complications with hospital costs and length of stay in a tertiary care center. J Gen Intern Med. 2006;21(2):177–80. Epub 2006/04/12. JGI319 [pii] pmid:16606377

[2] Hobson C, Ozrazgat-Baslanti T, Kuxhausen A, Thottakkara P, Efron PA, Moore FA, et al. Cost and Mortality Associated With Postoperative Acute Kidney Injury. Ann Surg. 2014. Epub 2014/06/03.

[3] Korenkevych D, Ozrazgat-Baslanti T, Thottakkara P, Hobson CE, Pardalos P, Momcilovic P, et al. The Pattern of Longitudinal Change in Serum Creatinine and 90-Day Mortality After Major Surgery. Ann Surg. 2015.

[4] Bihorac A, Brennan M, Ozrazgat-Baslanti T, Bozorgmehri S, Efron PA, Moore FA, et al. National surgical quality improvement program underestimates the risk associated with mild and moderate postoperative acute kidney injury. Crit Care Med. 2013;41(11):2570–83. pmid:23928835

[5] Bihorac A, Yavas S, Subbiah S, Hobson CE, Schold JD, Gabrielli A, et al. Long-term risk of mortality and acute kidney injury during hospitalization after major surgery. Ann Surg. 2009;249(5):851–8. pmid:19387314

[6] Huber M, Ozrazgat-Baslanti T, Thottakkara P, Efron PA, Feezor R, Hobson C, et al. Mortality and Cost of Acute and Chronic Kidney Disease after Vascular Surgery. Annals of vascular surgery. 2015.

[7] Ozrazgat-Baslanti T, Thottakkara P, Huber M, Berg K, Gravenstein N, Tighe P, et al. Acute and chronic kidney disease and cardiovascular morltaity after major surgeyr Annals of Surgery. 2015.

[8] Zarjou A, Agarwal A. Sepsis and acute kidney injury. J Am Soc Nephrol. 2011;22(6):999–1006. Epub 2011/05/14. ASN.2010050484 [pii]. pmid:21566052

[9] Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg. 1991;78(3):355–60. Epub 1991/03/01. pmid:2021856

[10] Lake AP, Williams EG. ASA classification and perioperative variables: graded anaesthesia score? Br J Anaesth. 1997;78(2):228–9. Epub 1997/02/01.

[11] Hobson CE, Singhania G, Bihorac A. Acute kidney injury in the surgical patient. Crit Care Clin. 2015.

[12] Huen SC, Parikh CR. Predicting acute kidney injury after cardiac surgery: a systematic review. Ann Thorac Surg. 2012;93(1):337–47. S0003-4975(11)02155-2 [pii] pmid:22186469

[13] Miyagawa T, Nishida N, Ohashi J, Kimura R, Fujimoto A, Kawashima M, et al. Appropriate data cleaning methods for genome-wide association study. J Hum Genet. 2008;53(10):886–93. Epub 2008/08/13. pmid:18695938

[14] Osborne JW. Is data cleaning and the testing of assumptions relevant in the 21st century? Front Psychol. 2013;4:370. Epub 2013/06/28. pmid:23805118

[15] Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS Med. 2005;2(10):e267. Epub 2005/09/06. 04-PLME-PF-0256 [pii] pmid:16138788

[16] Paul Thottakkara, Tezcan Ozrazgat-Baslanti, Bradley B. Hupf, Parisa Rashidi, Panos Pardalos, Petar Momcilovic, Azra Bihorac, Application of Machine Learning Techniques to High-Dimensional Clinical Data to Forecast Postoperative Complications, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155705

[17] https://pubmed.ncbi.nlm.nih.gov/29323142/