IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Flight Safety Detection Using Modified Zeror Approach with Nominal Data Conversion

Aparpreet Singh[1*], Sandeep Sharma[2]

[1,2]Department of Computer Engineering & Technology, Guru Nanak Dev University, Amritsar, Punjab, India

[*]*Corresponding Author: aparpreet@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i3.882887 | Available online at: www.ijcseonline.org

Accepted: 12/Mar/2019, Published: 31/Mar/2019

Abstract— The flight safety monitoring becomes critical and is core area of research focused upon by this literature. To this end, data mining mechanisms employed by existing literature are discussed. ZeroR classifiers shortcoming of handling string values are overcome by converting the attributes to nominal form. Overall process of improving classification process is divided into phases. First phase includes loading the dataset. The fetched dataset requires storage. The dataset is stored within local storage. Second phase is critical and requires additional storage for maintaining pre-processed dataset. Pre-processed dataset contains nominal data. ZeroR cannot handle string data hence pre-processing phase converts data in understandable format for ZeroR classifier. In the second phase, necessary fields required for result prediction are retained and rest of the fields are ignored using regression mechanism. Third phase is a classification phase indicating that the performance is above baseline or not. By accommodating, nominal value conversion process within ZeroR classifier, classification accuracy is improved by 15%.

Keywords— ZeroR, Classification accuracy, Nominal values, String data

I. INTRODUCTION

Today in aviation field the growth of air traffic demanded and required effective management. Because of high traffic it is difficult to ensure the safety of passengers and cargo. To resolve this, data acquisition is done and sensors are installed on aircrafts to handle airport data. The data that are extracted is too big and it is difficult to handle this data with traditional computers with database capabilities. The big data is considered one solution [1] to handle aviation safety.

The organised and raw data both are stored in dataset, this complex and extensive data is known as Big Data. The data are collected from various sensors that captured atmospheric data, network locales that are web based and so on, which is known as Big Data. By using Data mining the valuable data is extracted from this Big Data and in data mining there are various techniques that are used to model this Big Data.

Proposed system of ZeroR with nominal conversion is divided into phases. These phases includes the steps listed as under-

1.1 Pre-processing

Pre-processing mechanism is used in order to reduce the noise if any within the dataset. Noise can be in terms of missing data or abnormal values. The proposed system employ nominal conversion mechanism that converts string values into real values so that data can be processed by the classifier. In addition, the missing values are replaced with the significant values. The purpose of this mechanism is to introduce improvement in the classification process.

1.2 Online Buffer management

This step is critical in order to process the data at the server. The pre-processed data must be loaded at server side in order to be processed by mining phase. Buffer management component checks for the size of the pre-processed dataset and loads it into the buffer if capacity allows.

1.3 Clustering

This component is critical that partitioned the pre-processed data into clusters of similar entities. The advantage of using this step is to enhance the execution speed of classification process.

1.4 Data Mining

Data mining mechanism in the form of ZeroR is applied. This is a regression based approach used on the principal of fitting a curve. The equation of the straight line is used in this formation.

$$y = mx + b$$

Equation 1: Straight line fitting mechanism

© 2019, IJCSE All Rights Reserved

The mechanism uses method of least square in order to determine the values of the unknown 'm' and 'b'. Equations corresponding to method of least square within ZeroR are given in Equation 2.

$$\sum y = nb + m\sum x$$

Equation 2: Single level least square equation for the evaluation of variables

At the second level of equation, ZeroR evaluate next level of equation as

$$\sum y^2 = b\sum x + m\sum x^2$$

Equation 3: Second level ZeroR equation

These equations evaluation result in the result in terms of variables 'm' and 'b'.
The evaluation process yields the classes that give whether flight data falls within desired limits of class or not.

The aim of this research is to identify the risk factors while analyzing big data corresponding to flights. The pre-processing mechanism used in this literature eliminates the infrequent values and hence classification accuracy improves as classification is being performed. The ZeroR with nominal conversion is used to tackle all data types within the given dataset. Result is obtained in terms of classification accuracy.

Rest of the paper is organized as under: section 2 gives the related work of the mining approaches used for predicting abnormalities, section 3 gives the proposed system and methodology, section 4 gives performance analysis and result, section 5 gives conclusion and last section presents references.

II. RELATED WORK

In this section various techniques that are utilized for flight safety mechanisms are discussed.
In the thorough review of various data mining based techniques [2]proposed a mechanism to analyse big data corresponding to flight safety monitoring. Architectural framework for flight safety monitoring using big data technology with high precision tackled through this literature. It is used [3] for analysis of the data of shifting degree of flight and then use various data mining techniques for evaluating this data. In this literature various clustering algorithms are also overviewed and these algorithms are based on cluster processing that inspects usage decisions. It is mainly based on two fundamental decisions that use separation metric in time span and this time arrangement is used for characterisation. The mean time based approach [4] is used that are used to develop prediction based on the fluctuation that are gathered from flight. Various researches used clustering based approaches to formulate decisions about flight safety. It uses separation metric that are created in particular time span and then gives "medoid" on the basis of agent focus. In this time based correlation and data mining techniques are used that render comparable arrangement for separating data. The DTW approach [5] conveys data about the closeness and move between two time arrangements, thus represents something like one huge reason for blunder in time arrangement data mining. In a review of time-arrangement clustering strategies, it detailed that the vast majority of the distinguished systems can be gathered by their utilization of the first data. After utilizing clustering and data mining system it gives various abnormal state that are specified in this dataset. The clustering mechanism is used for locating neighbourhood and it does not use any adjustment strategy. It also extract data from datasets and after that clustering is applied that separate the highlighted data from this dataset. It [6] proposed the model that uses parameters of flight data and then patterns, practices are used. Then again, the age of various models in subsets or cases of the data take into account the utilization of clustering on model parameters.
The comparative analysis of the existing literature is given in Table 1

Table 1: Comparison of various mechanisms to check the prediction accuracy

Author/Reference	Technique	Merit	Demerit
Zhu et al.	Time series analysis	Time series analysis can be used to predict the abnormal activities with high execution speed.	Pre-processing mechanism in this literature is not considered at all
Jasra et al.	Series deviation and fluctuations	Pre-processing mechanism is employed in order to introduce clarity within the series of data	Regression mechanism can be accommodated to improve classification accuracy
Li et al.	DTW	Better classification accuracy in applications other than	Yet to be tested in the field of aviation industry

© 2019, IJCSE All Rights Reserved

		aviation industry	
Sun et al.	Clustering KNN	KNN clustering efficiently locate neighbours and hence better accuracy and least execution time is observed	Noisy data can hamper the performance of clustering mechanism

III PROPOSED SYSTEM

The proposed framework will be demonstrated by following phases:-

3.1 Data acquisition
3.2 Pre processing with nominal data
3.3 Data extraction
3.4 Data mining
3.5 Data analysis
3.6 Data visualization

3.1 Data acquisition: It collects data from online sources and stores it into buffer in memory. The raw data about flight, audio and video data are collected. Data storage is on the basis of capacity of the buffer. The buffer maintained at the client end is 1TB hence large dataset can be accommodated within the buffer using acquisition phase.

If (Data<Threshold)
$Buffer = Data_i$
End of if

This pseudo code reflects the conditional storage of data within buffer in case size of data is less than the size of maintained buffer.

3.2 **Pre processing with nominal data**: In this phase preprocessing is done using modified nominal conversion that handles string data efficiently. Traditional mining approaches do not handle string type data and it is difficult to classify this data so in this phase we utilize modified converters. The data is fed into the pre-processing phase and first of all missing data is tackled. Once the missing data is tackled, nominal conversion converts the string type data into real values. The pseudo code for the same is given as under-

$for\ i = 1: size(Buffer)$
$if\ Buffer_i(Rows, Cols) == then$
$MPV_i = Calculate\ Mode\ of\ Buffer(Rows_i, Cols)$
$Buffer_i(Rows, Cols) = MPV_i$
$end\ of\ if$
$if\ isString\big(Buffer_i(Rows, Cols)\big)$
$Buffer_i(Rows, Cols) = Numel\big(Buffer_i(Rows, Cols)\big)$
$End\ of\ if$
$End\ of\ For$

3.3 Data extraction: In this phase the data that are required for analysis is kept and unwanted data are removed. The extraction of data is based on criticality of data. The critical information is identified on the basis of repetition factors present within the data. The measure of central tendency is used to obtain criticality of information.

$$\begin{bmatrix} Value & Dataset_i > Threshold \\ 0 & Otherwise \end{bmatrix}$$

3.4 Data mining: It is the process of making decision using the extracted data. In this phase the ZeroR classification is utilized that find and describes the data classes and concepts.

ZeroR classification:-
In this model the training set is analyzed to build classifier and then classification is done. In classification data are predicted by using class labels. These dataset are further divided into test set and training set. Further analysis is done using this test set and it randomly sampled dataset. The tuples remaining that are not used to build classifier are independent of training set and dataset. The classifiers accuracy is estimated using test set. It will give the test tuples that are classified correctly. It uses cross validation to predict the higher accuracy. The steps are given in the form of equations and pseudo code as

According to equation 2 and 3 the formation of data structure to evaluate unknown is given as under

$\sum y = nb + m\sum x$ ----------2

$\sum y^2 = b\sum x + m\sum x^2$----------3

The formed matrix for the evaluation of 'b' and 'm' is given as under

$$\begin{matrix} \sum y = & n & \sum x \\ \sum y^2 = & \sum x & \sum x^2 \end{matrix}$$

The evaluation yield values of m and b which will be substituted within equation 1 to obtain fitting class.

3.5 Data Analysis: To check the format of data this phase is utilized. It analyse the structure of data that are classified by the data mining phase for decision making.

3.6 Data Visualization: The final phase of the proposed system that provides overall results in terms of accuracy, specificity and sensitivity.

© 2019, IJCSE All Rights Reserved

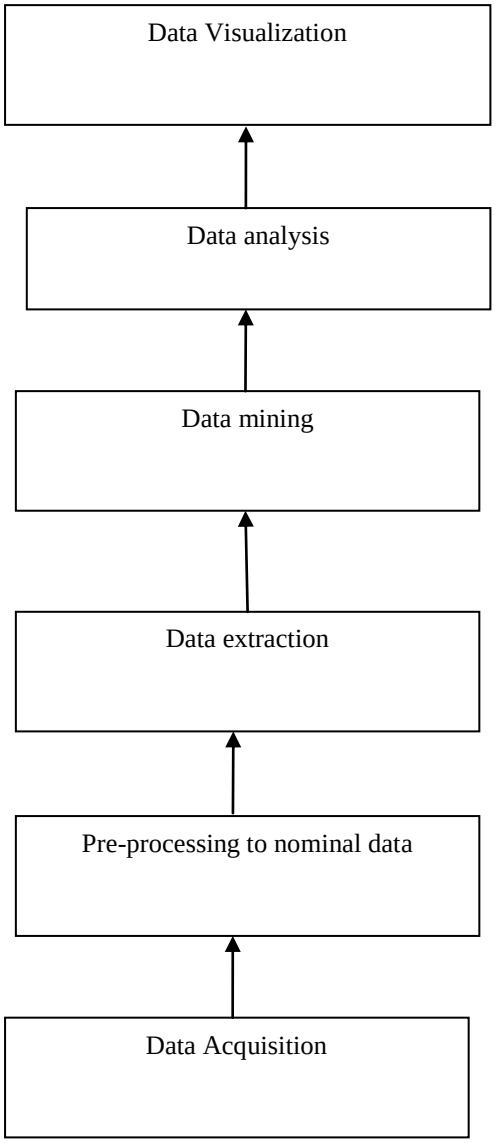

Fig 1: Proposed System Flow

The proposed system of ZeroR with Nominal conversion produces efficient result and shows improvement by significant margin. Next section present performance evaluation in terms of result generation phase.

IV RESULTS

The proposed system gives results in terms of classification accuracy, specificity and sensitivity. The results are as given below:

4.1 Accuracy

The accuracy is generated in terms of total true values generated from the dataset along with total negative values neglected by the proposed system.

Table 2: Classification accuracy comparison of ZeroR with Proposed system

Test data size	ZeroR system	ZeroR with Nominal Conversion system
2	74	83
5	76	84
10	78	86
15	73	81
20	79	87

This metric must be enhanced and indicates correctly identified segments from the records within the dataset. Classification accuracy is expressed in the form of true positive, true negative , false positive and false negative parameters as in the equation 4

$$Classification_{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 4: Classification accuracy evaluation

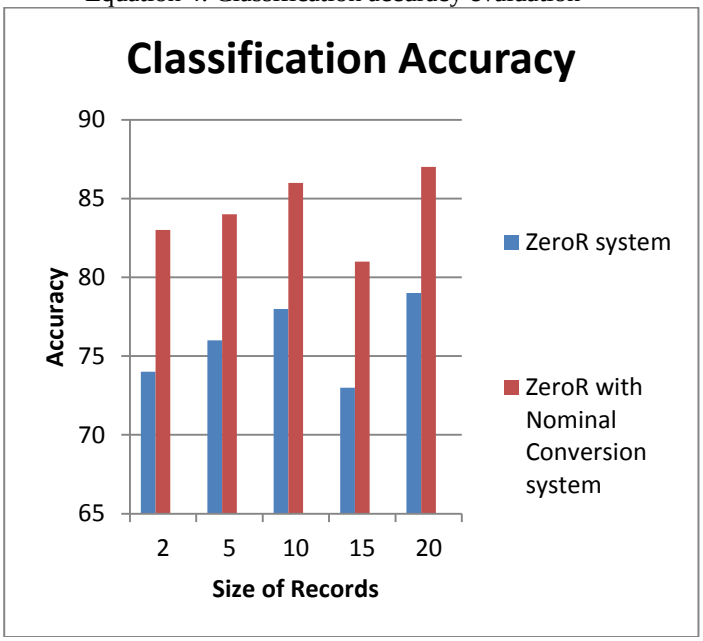

Figure 2: Comparative classification plot

4.2 Specificity

It is the metric used to determine total negative predicted to the total number of negative. Classification accuracy is hampered by this metric.

$$Specificity = \frac{TN}{N}$$

Equation 5: Specificity evaluation

© 2019, IJCSE All Rights Reserved

Results in terms of specificity is given in table 3

Table 3: Specificity through existing and proposed mechanisms

Test data size	ZeroR system	ZeroR with Nominal Conversion system
2	70	82
5	71	83
10	74	85
15	78	91
20	77	90

Specificity factor must be increased and is subsequently increased through the application of nominal conversion within ZeroR classifier. The pre-processing mechanism is the major contribution proposed through this literature and accuracy in terms of specificity is significantly improved as shown through the plot 3

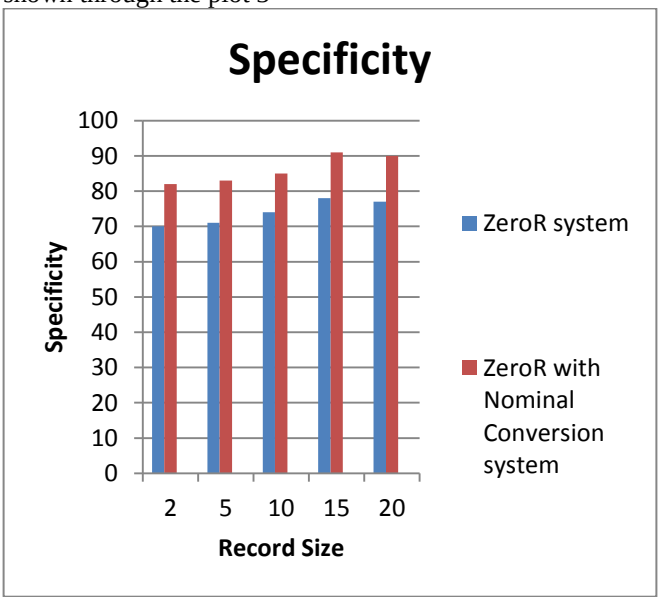

Figure 3: Specificity plots through ZeroR and ZeroR with Nominal conversion

4.3 Sensitivity

It is the metric used to determine the number of correct positive predictions to the number of positives. Classification accuracy is increased in case sensitivity is high.

$$Sensitivity = \frac{TP}{P}$$

Equation 6: Sensitivity evaluation

Sensitivity result is given in terms of classification accuracy through table 4

Table 4: Sensitivity Results with distinct data size

Test data size	ZeroR system	ZeroR with Nominal Conversion system
2	68	78
5	70	80
10	73	83
15	76	86
20	78	88

True positive rate indicates number of correct predictions through the proposed system. The TP rate is high using ZeroR but is further improved when pre-processing mechanism is employed and noise is terminated from the dataset. The plot 4 presents comparative distinctiveness of proposed and existing mechanism.

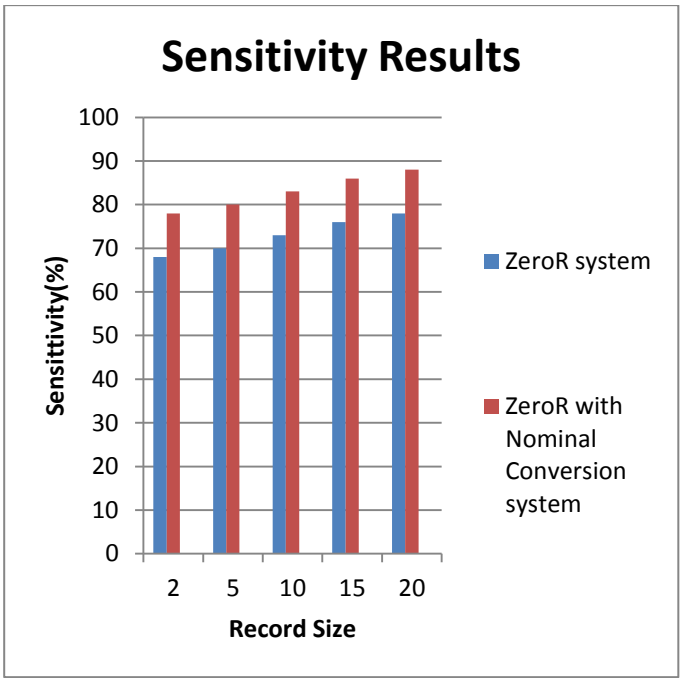

Figure 4: Sensitivity % comparison of ZeroR and ZeroR with Nominal conversion

V. CONCLUSION & FUTURE WORK

In this paper data mining approach based on ZeroR classifier is proposed that are utilized for flight safety monitoring and prediction. It demonstrates an application that uses high speed safety mechanism that observed airline data and make decisions. The result section shows that nominal value conversion process has high accuracy as compared to existing system. By using ZeroR classifier that handles string data efficiently and regression mechanism is utilized for better understanding. Future work ought to be done in breaking down the responsiveness and precision required to look after flying machine wellbeing. A more inside and out

© 2019, IJCSE All Rights Reserved

take a gander at the sensor information that have accessible in flight may give better expectations of airspeed.

REFERENCES

[1] C. Li, L. Zhu, and Z. Luo, "Big Time-frequency Domain Data Mining for Underdetermined BSS Using Density Component Analysis," *IEEE Access*, 2016.

[2] B. Li, X. Ming, and G. Li, "Big Data Analytics Platform for Flight Safety Monitoring," IEEE 2017 pp. 350–353, 2017.

[3] G. Zhu, K. Song, and P. Zhang, "A Travel Time Prediction Method for Urban Road Traffic Sensors Data," *2015 Int. Conf. Identification, Information, Knowl. Internet Things*, pp. 29–32, 2015.

[4] S. Jasra, J. Gauci, A. Muscat, and G. Valentino, "Literature review of machine learning techniques to analyse flight data," *Res. Gate*, no. October, 2018.

[5] G. Li, T. Yuan, S. J. Qin, and T. Chai, "Dynamic time warping based causality analysis for root-cause diagnosis of nonstationary fault processes," *Int. J. Autom. Control*, pp. 1289–1294, 2015.

[6] V. M. Janakiraman and D. Nielsen, "Anomaly Detection in Aviation Data using Extreme Learning Machines," 2016.

Authors Profile

Mr. Aparpreet Singh pursued Bachelor of Technology in Computer Science and Engineering from Guru Nanak Dev University, Amritsar.Currently, he is pursuing Master's of Technology in Computer Science and Engineering from Guru Nanak Dev University, Amritsar. His research interest is Big Data, Data Analytics and Data Mining.

Dr. Sandeep Sharma has done his B.E in Computer Science and Engineering, M.E in Computer Science and Engineering and Phd. His area of interest is Big Data,Cloud Computing and Parallel Procesing .Currently,he is Head and Professor at Department of Computer Engineering and Technology, Guru Nanak Dev University Amritsar.

© 2019, IJCSE All Rights Reserved