International Journal of Emerging Technology and Innovative Engineering
Volume 6, Issue 02, February 2020 (ISSN: 2394 – 6598)

MACHINE LEARNING BASED ANALYSIS FOR ROAD ACCIDENT PREDICTION

Arun Venkat[1], Gokulnath M[1], Guru Vijey K.P[1], Irish Susan Thomas[1], D.Ranjani[2]

1 Department of IT, Sri Krishna College of Technology, Coimbatore.

2 Assistant Professor, Department of IT, Sri Krishna College of Technology, Coimbatore.

ABSTRACT

Recent studies have predicted that in 2030, traffic accidents will be the fifth leading cause of death worldwide. The root cause of traffic accidents is hard to determine these days due to complex combination of characteristics like mental state of driver, road conditions, weather conditions, traffic, and violations of traffic rules to name a few. The costs of fatalities and driver injuries due to traffic accidents greatly affect the society. Application of machine learning techniques in the field of road accidents is gaining popularity these days. The deployment of machine learning classifiers has replaced traditional data mining techniques for generating higher results and accuracy. This work presents a survey of various existing work related to accident prediction using machine learning area.

KEYWORDS: Machine Learning, Accident Prediction, Survey, classification Techniques

1. INTRODUCTION

Road accident is one of the most prominent issues in the modern times of equipped and fast-moving traffic on roads. World Health Organization has reported top ten disastrous reasons for taking human's life, and unfortunately road accidents come at ninth place, where cardiac arrest sits on the top. The impact on society seems significant when cost of casualties and injuries from road accidents is evaluated. The young researchers these days have witnessed the increasing trend of evaluating the causes and implementing safety measures in order to preserve human life from dangerous road mishaps. Road traffic conditions and flow management continue to be an important area of research with many practical implications. During the last decade, the technological landscape of transportation has gradually integrated disruptive technology paradigms into current transportation management systems, leading to Intelligent Transportation Systems.

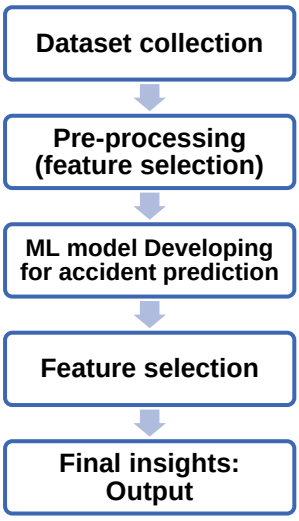


FIG.1. PROCESS INVOLVED IN ROAD ACCIDENT PREDICTION

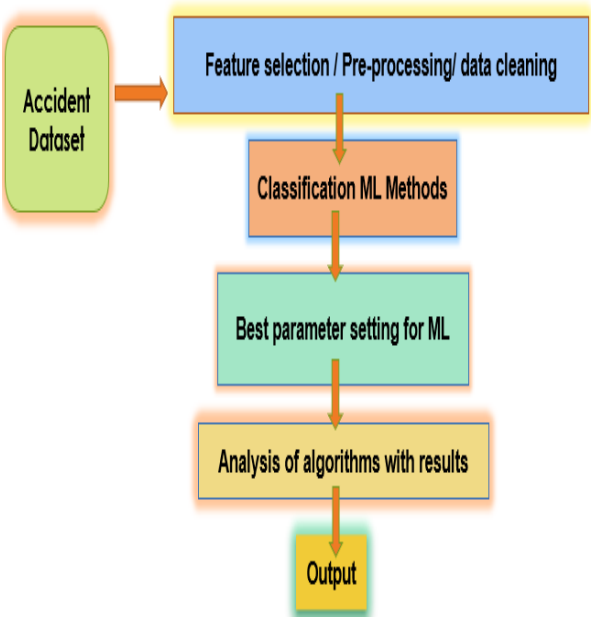
The major reason that data mining has attracted a great deal of attention in information industry in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. Classification algorithms are used to classify large volume of data and to provide interesting results. Application of data mining on social issues has been a popular technique recently. Fatal rates due to road accidents contribute more on the total death rate of the world. Over 1.2 million people die each year on the world's roads and between 20 and 50 million suffer non-fatal injuries. The report by says that around 1.2 million people were killed and 50 million injured in traffic collisions on the roads around the world each year and was the leading cause for death among children 10 – 19 years of age. Many road related factors which increase the death ratio were discussed in various literatures.

2. LITERATURE SURVEY

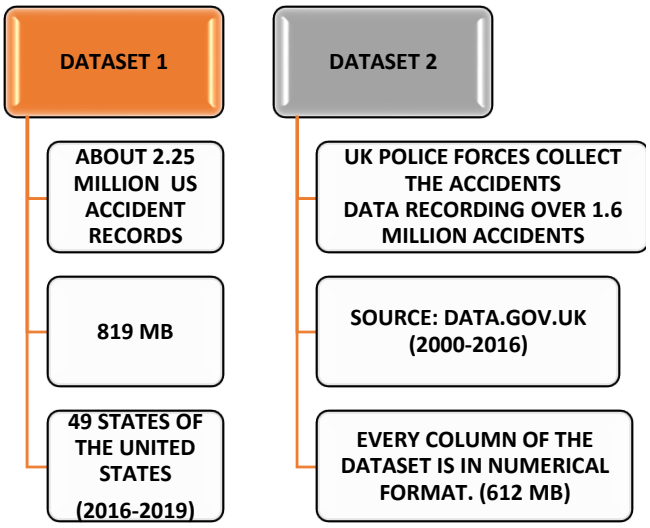
This section discusses various studies which have been conducted to emphasize the use of classification algorithms, feature relevance algorithms and Meta learners. Road accident is a global issue with skyrocketing trend where the immediate need is to analyse data through some data classification algorithms which help in anticipating the relevant factors influencing the problem of accident severity. Sobhan Sarkar et al designed a machine learning method using various algorithms. Two popular machine learning algorithms, namely support vector machine (SVM) and artificial neural network (ANN) have been used whose parameters are optimized by two powerful optimization algorithms, namely genetic algorithm (GA) and particle swarm optimization (PSO) in order to achieve higher degree of accuracy and robustness. PSO-based SVM outperforms the other algorithms with highest level of accuracy and robustness. Furthermore, rules are extracted by incorporating decision tree C5.0 algorithm with PSO-based SVM model. Finally, a set of nine useful rules extracted to identify the root causes behind the injury, near miss and property damage cases. A case study from a steel plant is presented to reveal the potentiality and validity of the proposed methodology. SVM, developed by Vapnik , is an emerging machine learning technique in statistical learning theory of multi-dimensional function which is used for classification and regression analysis. It holds an ability of being universal approximators of any multivariate functions to any desired level of accuracy. Initially, it was developed for regression tasks, but later was used as a powerful classifier. According to the previous studies, SVM has been used in most engineering fields with good accuracy. Theoretically, it has less overfitting problem, and better generalization ability. However, the main problem encountered in constructing SVM model is to adequately select training parameter values as inappropriate parameter setting leads to poor prediction accuracy. The readers may refer for basic understanding of the working principle of SVM. ANN is an artificial model of the human brain which can learn through adapting the present situations. It consists of interconnected network of neurons and synapses. Usually, it has three layers (i.e., input, hidden and output) or more (when more than one hidden layer). Hidden layers are considered the root of all calculations in ANN. A network gets activated when a set of inputs are triggered that consequently produce desired results through output layers. Each input value is multiplied by its corresponding weight layers, then it is summed up and added to a scalar parameter called bias, which in turn generates output through final output layer. Modifying connection weights and biases using appropriate learning algorithm, training process can be accomplished.

Many evolutionary algorithms or gradient descent methods have been used in this training process by updating weights and biases. At each iteration, they are modified until prediction error of the network gets minimized. Out of many learning algorithms, back propagation (BP), which is a gradient type of adjustment for the modification of weights, has been used in the paper of Benjio et al. Basically, the output of any node is determined by a mathematical operation on the input of the particular node. This operation is called the transfer function which facilitates the transformation of inputs into output either in linear or non-linear manner. There are three types of transfer functions used commonly in literature i.e., sigmoid, hyperbolic tangent, and linear. In this paper, sigmoid transfer function has been used. SVMs and artificial neural network (ANN) have shown better performance than other machine learning algorithms in some application areas, such as speech recognition, computer vision, and medical diagnosis. Although, SVM and ANN have an inherent inability to explain models and results, these algorithms construct black box models and learn patterns with no transparency and comprehensibility to humans. This drawback of these models impedes their application in some areas. Therefore, addressing this issue, in this study, we tried to extract meaningful rules from SVM to interpret the models. In literature, a proliferation of rule-extraction methods for trained SVMs has been proposed. Fu et al. has classified these motifs into three basic categories: “decomposition” (or transparent), “pedagogic” (or learning based), and “eclectic” (or hybrid). The transparent approach focuses on extracting region-based rules by support vectors (SVs) and separating hyperplane. For instance, Núñez et al. proposed the SVM and prototype method, and utilized the defined regions (ellipsoids and hyper rectangles) to refine the rules . Zhang et al. proposed the hyper rectangle rule extraction (HRE) algorithm, and Fung et al. suggested a linear programming formulation approach for rule extraction from linear SVMs. By contrast, the pedagogical approach treats SVMs model as a black box and uses the generated model to predict the label (class) for an extended data or unlabelled data. Barakat & Diederich used the resulting patterns to train a decision tree learning system and to extract the corresponding rule sets. The eclectic approach incorporates both “decomposition” and “pedagogic” techniques; it only uses the SVs or applied rule-based model to train the artificial data based on SVs. Barakat & Bradley proposed a SQRex-SVM algorithm based on the sequential covering approach. The proposed method extracts rules directly from the SVs of a trained SVM using a modified sequential covering algorithm. It was observed that the proposed method exhibited both improved generalization performance and smaller as well as comprehensible rule sets compared to both other SVM rule extraction techniques and direct rule learning techniques.

3. PROPOSED FRAMEWORK



3.1. Dataset



4. MACHINE LEARNING METHODS

a. Random Forest

Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. Random forest has nearly the same hyperparameters as a decision tree or a bagging classifier. Fortunately, there's no need to combine a decision tree with a bagging classifier because you can easily use the classifier-class of random forest. With random forest, you can also deal with regression tasks by using the algorithm's regressor. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

b. Logistic Regression

Logistic regression is another technique borrowed by machine learning from the field of statistics. It is the go-to method for binary classification problems (problems with two class values). In this post you will discover the logistic regression algorithm for machine learning. Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It’s an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

c. Decision Tree

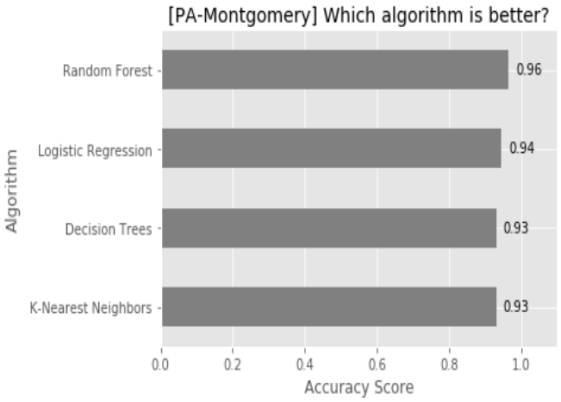
Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification **problems** too. The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by **learning decision rules** inferred from prior data (training data). The understanding level of Decision Trees algorithm is so easy compared with other classification algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each **internal node** of the tree corresponds to an attribute, and each **leaf node** corresponds to a class label.

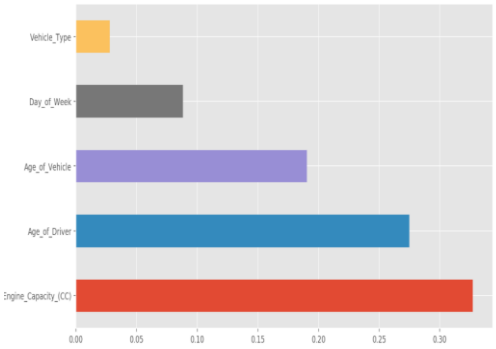
d. Nearest Neighbors Algorithm

The K-Nearest Neighbors algorithm (KNN) is a non-parametric method, which considers the K closest

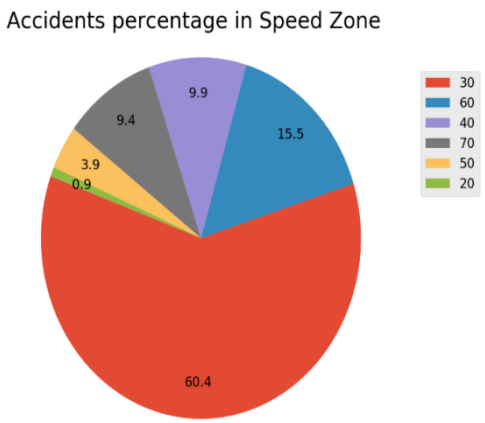
training examples to the point of interest for predicting its class. This is done by a simple majority vote over the K closest points. **KNN** algorithms use data and classify new data points based on similarity measures (e.g. distance function). **KNN** is a **supervised learning** algorithm used for classification.

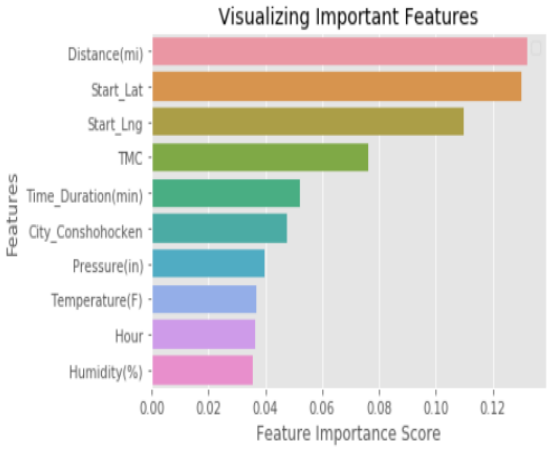
RESULTS FOR DATASET 1





RESULTS FOR DATASET 2





5. CONCLUSION

Recent studies have predicted that in 2030, traffic accidents will be the fifth leading cause of death worldwide. The costs of fatalities and driver injuries due to traffic accidents greatly affect the society. These insights call for investigating various aspects of traffic accident data analysis and modelling in numerous geographic regions. In particular, several researchers paid increasing attention to determining factors that greatly affect the severity of driver injuries caused by traffic accidents. This paper contributes various existing survey and works carried by researchers. The future scope of this work is to consider sentiment analysis of road accident's cases using ensemble classifiers and deep neural network.

6. REFERENCES

[1] Sánchez AS, Fernández PR, Lasheras FS, Juez FJDC, Nieto PJG. Prediction of work-related accidents according to working conditions using support vector machines. Appl Math Comput 2011;218:3539–52. doi:10.1016/j.amc.2011.08.100.

[2] ILO. Promoting safe and healthy jobs : The ILO Global Programme on Safety , Health and the Environment (Safework). 2008.

[3] EUROSTAT. Labour Force Survey 2007 ad hoc module on accidents at work and work related health problems. In: European Communities: 2009.

[4] ILO. Safety in Numbers, Global Safety Culture at Work (The International Labour Organisation, Geneva). 2003.

[5] Khanzode V V, Maiti J, Ray PK. Occupational injury and accident research : A comprehensive review. Saf Sci 2012;50:1355--1367. doi:10.1016/j.ssci.2011.12.015.

[6] Kunce JT. Vocational interests and accident proneness. J Appl Psychol 1967;51:223–5.

[7] RL Babu, S Vijayan,(2016) "Wrapper based feature selection in semantic medical information retrieval", Journal of Medical Imaging and Health Informatics, 6 (3), pp.802-805.

[8] RA Suji, DB Anand, RL Babu,(2019) "An Efficient Image Enhancement Technique using Stationary Wavelet Transform Based Image Fusion", International Journal of Recent Technology and Engineering 8 (1), pp.2277-3878.

[9] RA Suji, DB Anand, RL Babu,(2019) "Polar ice image segmentation using improved estimation and normalization of illumination", Cluster Computing 22 (1), pp.S3811–S3819.

[10] RL Babu, S Vijayan,(2013) "SEMANTIC BASED INFORMATION RETRIEVAL FOR WEB PAGES", Journal of Theoretical & Applied Information Technology 54 (1)

[11] S.Anbu Karuppusamy "Survey on Cluster Based Energy Efficient Routing Protocol using Ant Colony Optimization in Mantes" Cikitusi Journal for Multidisciplinary Research, ISSN No: 0975- 6876, Volume 6, Issue-2, Feb 2019.

[12] S.Anbu Karuppusamy "Study on Intelligent Naive Bayesian Probabilistic Estimation Practice for Traffic Flow to Form Stable Clustering In VANET" International Journal Of Information and Computing Science, ISSN No: 0975-6876 Volume69, Issue-2, Feb 2019.

[13] S.Anbu Karuppusamy "Time Efficient Data Collection with Mobile Sink and Vmimo Technique in Wireless Sensor Networks" International Journal of Research, ISSN No: 2236-6124 Volume 8, Issue-4, Feb 2019.

[14] S.Anbu Karuppusamy "Energy Efficient Protocol for WSN Using Adaptive Predictor with Kalman Filter" Pramana Research Journal, ISSN No: 2249-2976, Volume 9, Issue-3, 2019.

[15] S.Anbu Karuppusamy "Some Investigations On Cross Layer Optimization Reliable End to End QOS Support for HMAC in WSN" Suraj Punj Journal for Multidisciplinary Research, ISSN No: 2394-2886, Volume 9, Issue-3, 2019.

[16] S.Anbu Karuppusamy "Some Studies on Dynamic Source Routing in Mobile Ad Hoc Network" Journal of Applied Science and Computations ISSN No: 1076-5131, Volume 6, Issue-3, March 2019

[17] S.Anbu Karuppusamy "Some Smart Network for Vehicle Speed Control System" Universal Review, ISSN: 2277-2723, Volume-8, Issue-3, Jan 2019

[18] S.Anbu Karuppusamy "Some Investigation Study on Challenging Security Issues in Manet" International Journal of Scientific Research and Review, ISSN: 2279-543X, Volume 8, Issue-3, 2019.

[19] S.Anbu Karuppusamy "Some Investigations on Relay Beam forming in MIMO Systems of 4G LTE Networks" International Journal of Management, Technology and Engineering ISSN: 2249-7455, Volume 9, Issue-3, March 2019.

[20] S.Anbu Karuppusamy "Smart IoT Automation Network System Using Renewable Energy" International Journal of Scientific Research and Review, ISSN: 2279-543X, Volume 8, Issue-3, 2019.

[21] S.Anbu Karuppusamy "Effective Implementation of Smart Meter Monitoring Through IOT" International Journal of Current Engineering and Scientific Research, - ISSN (Print): 2393-8374, (Online): 2394-0697, Volume-5, Issue-4, 2018

[22] S.Anbu Karuppusamy "Performance Investigation of TCP signal strength in Multihop Adhoc Network" Asian Journal of Research in Social Science and Humanities- ISSN 2249-7315, Vol.6 Sep 2016

[23] S.Anbu Karuppusamy "Performance Investigation on Adaptive Wiseman Protocol For Ad Hoc & Wireless Sensor Network" Asian Journal of Information Technology- ISSN 1682-3915, Vol.9 Sep 2016

[24] S.Anbu Karuppusamy "Certain Investigations on Hybrid N.W Gamma scheme for Improving Performance of fading channels in MANETs" International Journal of Scientific Research and Review, ISSN: 2279-543X, Volume 8, Issue-6, 2019.

[25] S.Anbu Karuppusamy "Effective Implementation of QPSK Based Grover Searching Method for Enhancing Fading Channel Performance in MANETS" Universal Review, ISSN: 2277-2723, Volume 8 Issue 6- Jan 2019

[26] S.Anbu Karuppusamy "Automatic Human Activity Recognition in Video Surveillance System using Versatile Quadric Activity Portion Classification Method" Journal of Medical Imaging and Health Informatics**,** EISSN: 2156-7026 (Online) Volume.9 No7 Sep2019

[27] S.Anbu Karuppusamy "Video Surveillance System against Anti-Terrorism by Using Adaptive Linear Activity Classification (ALAC) Technique" Journal of Medical Systems**,** ISSN: 0148-5598 (Print) 1573-689X (Online), Volume 3 Number 8 Aug 2019.

[28] S.Anbu Karuppusamy "Classification of Chronic Obstructive Pulmonary Disease (COPD) using Gabor Filter With SVM Classifier" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019.

[29] S.Anbu Karuppusamy "Classification of Chronic Obstructive Pulmonary Disease (COPD) using Regression with Gabor Filtrations and Random Forest Classification" International Journal of Advanced Trends in Computer Science and Engineering ISSN: 2249 – 8958, Volume-8 Issue-5, October 2019

[30] S.Anbu Karuppusamy "Improving the Performance of TCP in Ad-Hoc Networks based on signal Strength and buffering system" Journal of Applied Sciences Research, Vol. 8, No. 5, EISSN 1819-544X, May 2012.

[31] RA Suji, DB Anand, RL Babu,(2019) "An Efficient Image Enhancement Technique using Stationary Wavelet Transform Based Image Fusion", International Journal of Recent Technology and Engineering 8 (1), pp.2277-3878.

[32] RA Suji, DB Anand, RL Babu,(2019) "Polar ice image segmentation using improved estimation and normalization of illumination", Cluster Computing 22 (1), pp.S3811–S3819.

[33] Adaline Suji, R, Saravanan, R & Bright Anand, D 2017, 'ImageFusion-Novel Approach of Contrast Enhancement by Discrete Shearlet Transform (DST)',Journal of Computational and Theoretical Nanoscience, vol. 14, no. 3, pp. 1282-1287

[34] Bright Anand, D, Saravanan, R & Adaline Suji, R 2017, 'Adaptive Batch Mode Active Learning Technique Usingwith an Improved Time Adaptive Support Vector Machine for Classification of Remote Sensing Applications', Journal of Computational and Theoretical Nanoscience, vol. 14, no. 2, pp. 1108-1113.

[35] Adaline Suji, R, Saravanan, R & Bright Anand (2015)"A Contrast Based Image Fusion Technique Using Digital Shearlet Transform and Root Mean Square Contrast"- Middle East Journal of Scientific Research, volume 23,Issue 7,pages 1444- 1449.