e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

LEVERAGING MACHINE LEARNING FOR ROAD ACCIDENT PREDICTION AND PREVENTION: A COMPREHENSIVE REVIEW

Ranjeet Yadav*1, Dakshita Joshi*2

*1M.Tech Scholar, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India.

*2Assistant Professors, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India.

DOI: https://www.doi.org/10.56726/IRJMETS65859

ABSTRACT

The auto industry has made efforts to create safer vehicles, but traffic accidents are inevitable, as this review paper explains. If we develop accurate prediction models that are able to automatically classify the type of injury and severity of various traffic accidents, we might be able to identify patterns in dangerous crashes. These social and street mishap examples can be valuable to foster traffic security control approaches. We believe that measures should be based on scientific and objective surveys of the causes of accidents and the severity of injuries in order to achieve the greatest possible accident reduction effects with limited budgetary resources. The results of four machine learning paradigms used to model the severity of injuries sustained in traffic accidents are summarized in this paper. We considered brain networks prepared utilizing cross breed learning draws near, support vector machines, choice trees and a simultaneous half and half model including choice trees and brain organizations. The results of the experiment show that the hybrid decision tree-neural network approach performed better than the other individual machine learning paradigms.

Keywords: Safer Vehicles, Traffic Accidents, Accurate Prediction, Identify Patterns, Machine Learning.

I. INTRODUCTION

In an era characterized by rapid urbanization, increasing vehicular population, and burgeoning technological advancements, road accidents have emerged as a pervasive global challenge, transcending geographical boundaries and socio-economic disparities [1]. The human and economic toll of road accidents is staggering, with millions of lives lost annually, leaving indelible scars on communities and economies [2]. As societies grapple with the multifaceted complexities of modern transportation systems, there is an imperative to harness innovative approaches for understanding, mitigating, and preventing road accidents [3].

Traditional methodologies for road accident analysis, rooted in statistical models and deterministic approaches, have been the cornerstone of traffic safety research for decades [4]. While these methods have provided valuable insights, the evolving nature of road dynamics, the influx of diverse data sources, and the complexity of human behaviors behind the wheel necessitate a paradigm shift in analytical approaches [5]. Enter machine learning (ML), a transformative branch of artificial intelligence that empowers systems to learn patterns, extract insights, and make predictions from data [6].

The integration of machine learning into road accident analysis marks a paradigmatic shift, offering a dynamic and data-driven lens to decipher the intricate web of factors contributing to accidents [7]. This synthesis of advanced analytics, vast datasets, and algorithmic prowess promises to revolutionize our understanding of road safety dynamics. This review embarks on a comprehensive exploration of the landscape where machine learning intersects with road accident analysis, unraveling the state-of-the-art methodologies, challenges, and transformative potential [8].

A. Contextualizing the Problem:

The magnitude of the road accident predicament demands nuanced approaches that transcend traditional methodologies. The World Health Organization (WHO) reports that road traffic injuries account for approximately 1.35 million deaths annually, making them the leading cause of death among young people aged 5–29 years. Beyond the human toll, road accidents incur significant economic costs, straining healthcare systems, and hindering sustainable development. As countries strive to meet the Sustainable Development Goals (SDGs), addressing road safety emerges as an integral component, underscoring the urgency of innovative and effective solutions [9].

In this context, machine learning offers a transformative lens to scrutinize the intricate patterns and causative factors underlying road accidents. Unlike conventional methods that often struggle to handle the sheer volume and diversity of contemporary data, machine learning algorithms excel in processing and extracting meaningful insights from complex datasets [10]. These algorithms, ranging from traditional classifiers to sophisticated deep learning models, have the capacity to discern latent patterns, correlations, and nonlinear relationships within the plethora of variables influencing road safety.

B. The Rise of Machine Learning in Road Safety:

The convergence of machine learning with road safety research has witnessed a surge in scholarly interest and practical applications [11]. Researchers, engineers, and policymakers are increasingly recognizing the potential of ML to not only analyze historical accident data but also to predict and prevent accidents in real-time. This transition from a reactive to a proactive approach holds promise for ushering in a new era of road safety, wherein insights derived from ML models guide interventions, policies, and infrastructure development [12].

Machine learning applications in road safety are diverse, encompassing various facets of accident analysis, risk prediction, and human behavior modeling. One prominent avenue involves the utilization of ML algorithms to analyze historical accident data, discern patterns, and identify high-risk zones or temporal trends [13]. By uncovering hidden correlations, these models contribute valuable insights for designing targeted interventions and resource allocation.

Moreover, machine learning facilitates the development of predictive models that can forecast potential accidents based on real-time data inputs [14]. These models leverage information from sensors, traffic cameras, weather conditions, and other dynamic variables to create a predictive framework capable of alerting authorities and drivers about impending risks. The real-time nature of these predictions empowers stakeholders to implement preventive measures promptly, mitigating the likelihood of accidents [15].

C. Objectives of the Review:

Amidst the burgeoning literature on machine learning applications in road safety, this comprehensive review aims to achieve several key objectives:

Survey Existing Methodologies: Conduct an in-depth examination of the various machine learning methodologies employed in road accident analysis, ranging from classical models to state-of-the-art deep learning approaches. This survey will provide a nuanced understanding of the diverse tools available for researchers and practitioners.

Evaluate Performance Metrics: Critically assess the performance metrics used to gauge the effectiveness of machine learning models in predicting, preventing, and analyzing road accidents. By scrutinizing the strengths and limitations of existing metrics, this review aims to contribute to the ongoing discourse on benchmarking and evaluation standards.

Explore Challenges and Limitations: Identify and analyze the challenges and limitations associated with the application of machine learning in road safety. This exploration will shed light on the ethical, technical, and practical considerations that researchers and policymakers must navigate in harnessing the potential of ML for accident analysis.

Highlight Innovations and Future Trends: Illuminate innovative applications and emerging trends in the realm of machine learning for road safety. By identifying cutting-edge developments and foreseeing future trajectories, this review seeks to guide researchers, practitioners, and policymakers in shaping the trajectory of road safety research.

Synthesize Practical Implications: Distill practical insights and implications for real-world applications of machine learning in road accident analysis. By bridging the gap between theoretical advancements and on-the-ground implementations, this review aims to facilitate the translation of research findings into tangible interventions and policy frameworks.

In this review paper section I contains the introduction, section II contains the literature review details, section III contains the details about algorithms, section IV contains the software and language details, and section V provide conclusion of this review paper.

IRJMETS

e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

II. LITERATURE SURVEY

Yang and co. used a neural network approach to identify safer driving patterns that are less likely to result in injuries or deaths in car crashes [17]. In order to reduce the dimensions of the data, they carried out the Cramer's V Coefficient test [18] in order to locate significant variables that result in injury. After that, they used a frequency-based data transformation method to convert categorical codes into numerical values. Using a Backpropagation (BP) neural network, they made use of the University of Alabama-developed Critical Analysis Reporting Environment (CARE) system. They obtained a set of controllable cause variables that are likely causing the injury during a crash by utilizing the interstate alcohol-related data from 1997 Alabama and further investigating the weights on the trained network. There were two classes of the target variable in their study: injury and non-injury, where fatalities were included in the injury class. They discovered that they could potentially reduce fatalities and injuries by up to 40% by controlling a single variable, such as the driving speed or the lighting.

Sohn and co. used data fusion, ensemble, and clustering to boost the accuracy of individual classifiers for two categories of road traffic accident severity—physical injury and property damage—[15]. Neural network and decision tree classifiers were utilized as individual classifiers. After dividing the dataset into subsets with a clustering algorithm, they used each subset of data to train the classifiers. They discovered that when the variation in the observations is relatively large, as it is in the Korean data on road traffic accidents, clustering-based classification works better.

Mussone, others utilized neural networks to investigate a car accident that took place at an intersection in Milan, Italy [12]. BP learning-based feed-forward MLP was their choice. Eight variables—day or night, traffic flows circulating in the intersection, number of virtual conflict points, number of real conflict points, type of intersection, type of accident, condition of the road surface, and weather—had ten input nodes in the model. The ratio of the number of accidents at a given intersection to the number of accidents at the most dangerous intersection was used to calculate the output node, which was known as an accident index. According to the findings, the nighttime intersections with no traffic signals have the highest accident index for pedestrians being run over.

Dia and co. based a multi-layered MLP neural network freeway incident detection model on real-world data [5]. They thought about the presentation of the brain network model and the episode recognition model in procedure on Melbourne's expressways. The outcomes demonstrated that a neural network model could outperform the currently in use model in terms of incident detection speed and dependability. They also discovered that model performance in that section of the freeway could significantly suffer if speed data were not provided at a station.

Shankar and other used a nested logic formulation to estimate the likelihood of an accident's severity based on the likelihood of an accident happening [14]. They discovered that if at least one driver did not use a restraint system at the time of the accident, there is a greater chance of evident injury, disabling injury, or death than there is of no evident injury.

Kim et al. developed a log-linear model to explain how driver characteristics and actions contributed to more severe injuries. They discovered that driving under the influence of alcohol or drugs and not wearing a seat belt significantly raise the risk of more severe accidents and injuries [8].

Abdel-Aty and co used crash databases from the Fatality Analysis Reporting System (FARS) that covered the years 1975 to 2000 to look at how the rise in registrations for Light Truck Vehicles (LTV) affected fatal angle collision trends in the US [1]. They looked into the number of annual fatalities caused by angle collisions and the configuration of the collision (car-car, car-LTV, car-LTV, and LTV-car). The results of time series modeling indicated that fatalities as a result of angle collisions will rise over the next ten years, and that this rise will be influenced by the anticipated overall increase in the proportion of LTVs in traffic.

Bedard and co. utilized multivariate logistic regression to identify the independent contribution of driver, crash, and vehicle characteristics to the fatality risk of drivers [3]. They discovered that reducing speed, reducing the number and severity of driver-side impacts, and increasing seatbelt use may reduce fatalities. To ascertain the connection between accident notification times and fatalities, Evanco carried out a multivariate population-based statistical analysis [6]. The study found that the length of time it takes to notify drivers of an accident is a

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

significant factor in the number of fatalities resulting from collisions on rural roads. Ossiander and others utilized Poisson regression to examine the relationship between the speed limit increase and the fatal crash rate (fatal crashes per vehicle mile traveled) [13]. In Washington State, they discovered that an increase in the speed limit was linked to a higher rate of fatal crashes and an increase in fatalities on freeways.

Abdel-Aty and co used crash databases from the Fatality Analysis Reporting System (FARS) that covered the years 1975 to 2000 to look at how the rise in registrations for Light Truck Vehicles (LTV) affected fatal angle collision trends in the US [1]. They looked into the number of annual fatalities caused by angle collisions and the configuration of the collision (car-car, car-LTV, car-LTV, and LTV-car). The results of time series modeling indicated that fatalities as a result of angle collisions will rise over the next ten years, and that this rise will be influenced by the anticipated overall increase in the proportion of LTVs in traffic.

Bedard and co. utilized multivariate logistic regression to identify the independent contribution of driver, crash, and vehicle characteristics to the fatality risk of drivers [3]. They discovered that reducing speed, reducing the number and severity of driver-side impacts, and increasing seatbelt use may reduce fatalities. To ascertain the connection between accident notification times and fatalities, Evanco carried out a multivariate population-based statistical analysis [6]. The study found that the length of time it takes to notify drivers of an accident is a significant factor in the number of fatalities resulting from collisions on rural roads. Ossiander and others utilized Poisson regression to examine the relationship between the speed limit increase and the fatal crash rate (fatal crashes per vehicle mile traveled) [13]. In Washington State, they discovered that an increase in the speed limit was linked to a higher rate of fatal crashes and an increase in fatalities on freeways.

III. ALGORITHM

- **k-means clustering algorithm**

The well-known clustering problem can be solved with one of the simplest unsupervised learning algorithms, k-means. The method follows a straightforward method for classifying a given data set using a predetermined number of clusters (assume k clusters).

The primary objective is to identify k centers, one for each cluster. Because different locations result in different outcomes, these centers should be strategically placed. Therefore, placing them as far apart as possible is the best option.

The next step involves associating each point in a given data set with the closest center. The first step is finished and an early group age is completed when no point is pending. We need to recalculate k new centroids as the barycenter of the clusters from the previous step at this point.

A new binding needs to be made between the same data set points and the nearest new center once we have these k new centroids. There has been created a loop. Because of this circle we might see that the k communities change their area bit by bit until no more changes are finished or at the end of the day places move no more.

- **Logistic Regression**

Logistic regression is the regression analysis and dependent upon the variables is binary numbers i.e. (0s and 1s), All regression analysis, the logistic regression is a prediction analysis. Logistic regression is used to details about data and to graphically explain the relationship between dependent binary variable and more nominal, ordinal, interval independent variables.

Sometimes logistic regressions are difficult to describe the statistics tools are easily conduct and analysis the datasets, then in others plain word are as it is display in the output.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com

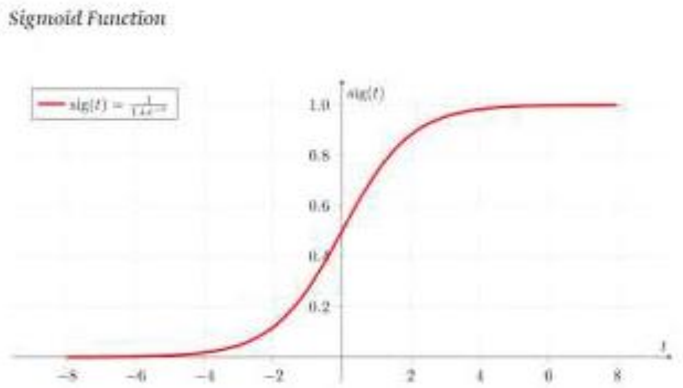

Figure 1: Signoid Function

IV. SOFTWARE AND LANGUAGES USED

- **Jupyter**

Jupyter's mission is the creation of open-source software. It is utilized in dozens of programming languages for open standards and interactive computing services. It is a live document and code creation and sharing open-source web application. Which is a significant benefit of Jupyter? Cleaning and transforming data, numerical simulation, statistical modeling, machine learning, and many other applications are all possible with it. The algorithm was run with Jupyter.

- **Python**

Python is a high-level, interpreted programming language with a wide range of applications. Python, developed by Guido van Rossum and first released in 1991, places an emphasis on code readability through the use of a lot of whitespace in its design. It is currently the programming language that is used the most. It offers structures that make it possible to program clearly at both small and large scales. The system's logistic regression is carried out using jupyter, and the algorithm is written in python.

- **HTML, CSS, and JSCRIPT**

The web development languages that are utilized the most frequently. These programming languages were used to create the prediction system's user interface. The website serves as an interface, passing the various constraints entered by users to the program for it to work with.

V. CONCLUSION

In concluding our comprehensive review of road accident analysis using machine learning, it becomes evident that the marriage of advanced analytical techniques with the complexities of road safety presents an exciting frontier for research, innovation, and practical applications. The synthesis of machine learning and road accident analysis offers a multifaceted approach to understanding, predicting, and preventing accidents, transcending the limitations of traditional methodologies.

The critical examination of existing literature showcased a diverse range of methodologies and models employed across studies. From predictive modeling using supervised learning algorithms to the exploration of patterns through unsupervised techniques, the field has witnessed an evolution marked by creativity and adaptability. This variability in approaches reflects the complex nature of road safety, where no singular solution can adequately address the myriad factors influencing accidents.

Technological interventions, such as the integration of the Internet of Things (IoT), computer vision, and sensor networks, have played a pivotal role in augmenting the capabilities of machine learning in road safety. These advancements not only enhance data collection but also empower real-time decision-making, creating a dynamic framework for accident prevention and mitigation. As the Fourth Industrial Revolution unfolds, the synergies between cutting-edge technologies and machine learning are likely to redefine the possibilities for improving road safety.

However, as we traverse this transformative landscape, it is crucial to acknowledge the challenges and ethical considerations associated with the integration of machine learning in road safety. Biases in algorithms, concerns over data privacy, and the potential for unintended consequences underscore the importance of responsible and equitable deployment of these technologies. Addressing these challenges is imperative for building trust, ensuring fairness, and maximizing the positive impact of machine learning in the pursuit of safer roads.

Looking ahead, the exploration of future directions in machine learning for road accident analysis reveals a horizon brimming with possibilities. Refining existing models, exploring novel applications, and fostering interdisciplinary collaborations are pathways toward continual advancement. The integration of machine learning with emerging technologies and the development of comprehensive, adaptive systems have the potential to revolutionize how we approach road safety.

In conclusion, the review underscores the transformative potential of machine learning in reshaping the road safety landscape. By synthesizing diverse research efforts, discussing technological interventions, and addressing challenges and ethical considerations, this review serves as a comprehensive resource for researchers, policymakers, and practitioners alike. As we collectively strive for a future where road accidents are mitigated and prevented, the integration of machine learning emerges as a powerful ally, guiding us toward safer and more efficient transportation systems on a global scale.

VI. REFERENCES

[1] Abdel-Aty, M., and Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. Accident Analysis and Prevention, 2003.

[2] Abdelwahab, H. T. and Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. Transportation Research Record 1746, Paper No. 01-2234.

[3] Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. Accident analysis and Prevention, Vol. 34, pp. 717-727, 2002.

[4] Buzeman, D. G., Viano, D. C., & Lovsund, P., Car Occupant Safety in Frontal Crashes: A Parameter Study of Vehicle Mass, Impact Speed, and Inherent Vehicle Protection. Accident Analysis and Prevention, Vol. 30, No. 6, pp. 713-722, 1998.

[5] Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. Transportation Research C, Vol. 5, No. 5, 1997, pp. 313-331.

[6] Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. Accident Analysis and Prevention, Vol. 31, 1999, pp. 455-462.

[7] Hand, D., Mannila, H., & Smyth, P., Principles of Data Mining. The MIT Press, 2001.

[8] Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. Accident Analysis and Prevention, Vol. 27, No. 4, 1995, pp. 469-481.

[9] Kweon, Y. J., & Kockelman, D. M., Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, 2003, pp. 441-450.

[10] Martin, P. G., Crandall, J. R., & Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp. 541-557.

[11] Mayhew, D. R., Ferguson, S. A., Desmond, K. J., & Simpson, G. M., Trends In Fatal Crashes Involving Female Drivers, 1975-1998. Accident Analysis and Prevention, Vol. 35, 2003, pp. 407-415.

[12] Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. Accident Analysis and Prevention, Vol. 31, 1999, pp. 705-718.

[13] Ossiander, E. M., & Cummings, P., Freeway speed limits and Traffic Fatalities in Washington State. Accident Analysis and Prevention, Vol. 34, 2002, pp. 13-18.

[14] Shankar, V., Mannering, F., & Barfield, W., Statistical Analysis of Accident Severity on Rural Freeways. Accident Analysis and Prevention, Vol. 28, No. 3, 1996, pp.391-401.

[15] Sohn, S. Y., & Lee, S. H., Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. Safety Science, Vol. 4, issue1, February 2003, pp. 1-14.

[16] Tavris, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. Accident Analysis and Prevention, Vol. 33, 2001, pp. 167-172.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:01/January-2025 Impact Factor- 8.187 www.irjmets.com
[17] Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns By A Neural Network Approach. ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining, Vol. 9, pp 839-844, Nov. 1999.

[18] Zembowicz, R. and Zytkow, J. M., 1996. From Contingency Tables to Various Forms of Knowledge in Database. Advances in knowledge Discovery and Data Mining, editors, Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. AAAI Press/The MIT Press, pp.329-349.

[19] Abraham, A., Meta-Learning Evolutionary Artificial Neural Networks, Neurocomputing Journal, Elsevier Science, Netherlands, Vol. 56c, pp. 1-38, 2004.

[20] Moller, A.F., A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Volume (6), pp. 525-533, 1993.

[21] National Center for Statistics and Analysis http://www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/NASS.html

[22] Roh J.W., Bessler D.A. and Gilbert R.F., Traffic fatalities, Peltzman's model, and directed graphs, Accident Analysis & Prevention, Volume 31, Issues 1-2, pp. 55-61, 1998.

[23] Peltzman, S., The effects of automobile safety regulation. Journal of Political Economy 83, pp. 677–725, 1975.

[24] Ossenbruggen, P.J., Pendharkar, J. and Ivan, J., Roadway safety in rural and small urbanized areas. Accid. Anal. Prev. 33 4, pp. 485–498, 2001.

[25] Abdalla, I.M., Robert, R., Derek, B. and McGuicagan, D.R.D., An investigation into the relationships between area social characteristics and road accident casualties. Accid. Anal. Prev. 29 5, pp. 583–593, 1997.

[26] Miaou, S.P. and Harry, L., Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 6, pp. 689–709, 1993.

[27] SVMlight. http://www.cs.cornell.edu/People/tj/svm_light/. Access date: May, 2003.

[28] Vapnik, V. N., The Nature of Statistical Learning Theory. Springer, 1995.

[29] Chong M., Abraham A., Paprzycki M., Traffic Accident Data Mining Using Machine Learning Paradigms, Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, ISBN 9637154302, pp. 415- 420, 2004.

[30] Chong M., Abraham A., Paprzycki M., Traffic Accident Analysis Using Decision Trees and Neural Networks, IADIS International Conference on Applied Computing, Portugal, IADIS Press, Nuno Guimarães and Pedro Isaías (Eds.), ISBN: 9729894736, Volume 2, pp. 39-42, 2004.

[31] Eui-Hong (Sam) Han, Shashi Shekhar, Vipin Kumar, M. Ganesh, Jaideep Srivastava, Search Framework for Mining Classification Decision Trees, 1996. umn.edu/dept/users/kumar/dmclass.ps

[32] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.

[33] Abraham, Intelligent Systems: Architectures and Perspectives, Recent Advances in Intelligent Paradigms and Applications, Abraham A., Jain L. and Kacprzyk J. (Eds.), Studies in Fuzziness and Soft Computing, Springer Verlag Germany, Chapter 1, pp. 1-35, 2002.