Course: MSc Data Analytics

Module Title: B9DA110 Advanced Data and Network Mining

Lecturer Name: Terri Hoare

Research Topic: " Predicting stress levels of automobile drivers using classical machine learning classifiers "

Context:

This paper explores the topic of stress levels in car drivers, specifically in the context of traffic congestion and other factors that can induce stress while driving. The main motivation is to understand the causes of stress in drivers, how it varies with different features like gender, age, driving skills, distraction, and driving concentration, and whether machine learning algorithms can effectively predict stress levels.

- The main issues and themes in this topic include the impact of stress on driving safety and overall well-being, the factors contributing to driver stress, and the potential for technology, such as wearable devices, to monitor and manage stress in real-time.
- The paper references relevant prior academic work, as it discusses various studies that have used different machine learning algorithms to predict stress levels in drivers. It also cites statistics from the World Health Organization (WHO) regarding road safety and the impact of road traffic crashes.
- Here, we know that Machine learning algorithms like K-Nearest Neighbors, Decision Trees, Naive Bayes, and PART have been applied to predict stress levels in car drivers. They look at data from wearables, like smartwatches, to make these guesses. Studies tell us that things like being a man or a woman, how old you are, and how focused you are while driving can make you more or less stressed when driving.
  The things we are not aware of are exactly which body signals show if a driver is stressed. We don't fully know how things like age or distraction affect stress in drivers. Different computer programs like Random Forest need testing to see if they're good at this. There might be other things we haven't checked that make drivers stressed, and we're not sure if this information works the same way everywhere.

Gap:

The paper identifies a gap in the literature related to monitoring and predicting stress levels in car drivers, particularly in the context of traffic congestion. While there is a body of research on this topic, the paper seeks to contribute by comparing the performance of various machine learning algorithms and exploring the impact of different demographic and situational factors on stress levels in drivers. It extends the envelope by providing insights into the effectiveness of different algorithms, gender and age-related differences, the influence of distraction and driving concentration, and potential improvements in classification performance through oversampling.

Question:

The research question addressed in this paper is whether machine learning algorithms are effective and appropriate tools for predicting stress levels in car drivers. The authors have indeed rephrased themes as questions, such as "Are machine learning algorithms effective and appropriate tools to predict stress levels?" This question is supported by sub-questions that relate to the impact of demographic factors, distraction, and driving concentration on stress levels in drivers.

How did they do the work?

The authors used an empirical method and conducted experiments involving 12 drivers who travelled different routes while wearing a biomedical device to collect physiological data. They applied four different data mining algorithms: K-Nearest Neighbor (KNN), Decision Tree, PART, and Naive Bayes. The data collection process involved physiological measurements, self-reported data, and environmental information, including road images, GPS location, and car speed. They used a publicly available implementation of the chosen classifiers and employed a standard ten stratified folds cross-validation approach for evaluation.

What did they find?

The paper presents the results of their experiments, which show that Naive Bayes outperformed other classifiers in terms of AUC and Precision, while KNN had better Recall and F1 scores as shown in the figure below. The results also indicate that stress levels differ based on gender, age, driving skills, distraction, and driving concentration.

Classifiers	Class	P	R	F1	Acc.	AUC
Naïve Bayes	non-stress	0.991	0.792	0.880	0.792	0.854
	stress	0.126	0.799	0.218		0.854
PART	non-stress	0.964	1.000	0.982	0.964	0.500
	stress	0.000	0.000	0.000		0.500
KNN	non-stress	0.972	0.996	0.984	0.968	0.848
	stress	0.674	0.227	0.340		0.848
Decision Tree	non-stress	0.964	1.000	0.982	0.964	0.500
	stress	0.000	0.000	0.000		0.500

(figure taken from IEEE research paper 2022)

What is the answer?

The analysis and discussion suggest that machine learning algorithms, particularly Naive Bayes and KNN, are effective for identifying stress in drivers. The data imply that physiological indicators can help predict stress levels while driving, and that these levels are influenced by various factors, as mentioned earlier. The paper has validated its results through the evaluation measures used in the experiments.

Significance:

The paper's significance lies in its contribution to understanding and predicting stress in drivers, which has implications for road safety and public health. It provides insights into the potential of using machine learning and physiological data to monitor and manage stress levels during driving. The contribution is original in the context of comparing multiple algorithms and considering various influencing factors. Its value extends to both academia and practice, as it adds to the knowledge base and offers practical implications for stress management during daily commutes. The impact of the paper on academia can be assessed by examining factors such as the number of citations and the quality of the journal in which it was published.

Studied Research paper details:

Published in: 2022 International Conference on Business Analytics for Technology and Security (ICBATS)
Date of Conference: 16-17 February 2022
Date Added to IEEE *Xplore*: 27 April 2022
INSPEC Accession Number: 21760729
DOI: 10.1109/ICBATS54253.2022.9759005
Publisher: IEEE
Conference Location: Dubai, United Arab Emirates