ABSTRACT

PREDICTIVE ANALYTICS WITH CLASS IMBALANCE: IN PERSONALIZED TRAVEL RECOMMENDATION AND AUTOMOTIVE CRASH INJURY RESEARCH

Paromita Shithi Nitu, M.S.
Marquette University, 2023

In recent years of information abundance, most real-world dataset encounters imbalance data distribution. Thus, machine learning and data mining predictive model requires careful adaptation for domain specific class imbalance situation. In this dissertation, an extensive search for ensemble algorithm modification has been performed to obtain thoracolumbar spine fracture (TL-fx) feature space since it shows an increasing incidence rate as a function of vehicle model year in automotive crash injury data base. Crash injury data has unique multilevel replication and nested pattern and no machine learning model performed adequately for class imbalance TL-fx. The aim of this dissertation is to develop a predictive algorithm modification (alternatively computational pipeline) to predict one or multiple TL-fx to understand the combine effect of occupant’s demographic features, vehicle condition and crash event along with range of associated injuries. To obtain the features associated with thoracolumbar spine injury, this dissertation proposes a brand new algorithm in combination of multiple modifications in different steps named as Garden algorithm primarily focused on achieving higher sensitivity as a goodness of fit and the underlining predictive classifier for this proposed modification is well-known random forest algorithm. The performance accuracy of Garden algorithm is ~71% which is ~8.7% higher than the mean accuracy of next best performing state of art classification model.

The extensive amount of information accessibility on electronic platforms (such as social media), urges the rapid development of the system that filters out irrelevant information and provides effective content that meet user-specific needs and expectations. In addition, this dissertation also investigates data imbalance in text classification to extract the travel preference based on user’s social media data and endeavors to implement proposed algorithm modification (alternatively text blocking) to predict user’s travel preference with skewed text category. A system architecture has been designed to obtain personalized travel recommendations and generate place of interest recommendation in a timely manner. The proposed model considers a user’s most recent interest by incorporating time-sensitive recency weight into the model thus, outperformed the existing personalized place of interest recommendation model and the overall accuracy is 75.23%.