International Journal of Information Technologies and Systems Approach
Volume 17 • Issue 1 • January-December 2024

Enhancing Car Segmentation for Thailand's Expressway Industry With an Automated Hybrid Machine Learning Framework

Kulkatechol Kanokngamwitroj
https://orcid.org/0000-0002-1734-0103
Rangsit University, Thailand

Chetneti Srisa-An
Rangsit University, Thailand

ABSTRACT

Car segmentation on Thailand's expressways poses challenges for traditional models due to unique characteristics, often resulting in predictive inaccuracies. Manual data analytics in this field is time-consuming and human centric. This research introduces an Automated Hybrid Machine Learning (AHML) framework leveraging advancements in AutoML, tailored for personalized customer segmentation in Thailand's expressway industry. The framework streamlines and automates the machine learning process, aiming to expedite model construction while enhancing performance. By employing clustering as an initial step followed by the Random Forest classifier as a hybrid classification approach, significant performance improvements are achieved compared to existing methods. Specifically, the model outperforms by 9.15% and 12.84% in both clusters, respectively. This research highlights the potential of the framework to address complex segmentation challenges and advance personalized customer targeting.

KEYWORDS

Machine Learning, Clustering, Classification, Automated Hybrid Machine Learning Framework

INTRODUCTION

The expressway industry has evolved into a universally recognized standard for transportation services, particularly in countries like Thailand. In Thailand, transportation by car has emerged as the predominant mode of travel. The extensive use of cars on expressways has made the development of cars a central pillar of economic growth strategies for Thailand and various developing nations over the years. This industry has played a pivotal role in generating substantial revenue in Thailand, amounting to billions of dollars. With millions of daily transactions occurring throughout Thailand, the transactions database has become the repository for an immense volume of both structured and unstructured data, resulting in what is commonly referred to as big data. The recognition of data's significance in understanding customer preferences and requirements has led to industry executives adopting data analytics as a crucial tool. This data-driven approach empowers businesses by giving them valuable insights into customer behaviors and needs. To further enhance the customer relationship model, this model enables an expressway organization to effectively monitor and respond to customer demands, ensuring high satisfaction. Consequently, to meet its customers' evolving expectations, the

DOI: 10.4018/IJITSA.353439

Expressway Organization of Thailand has a plan to launch a personalized customer service program, tailoring the initiative to address individual customers' unique needs and elevate the customer service of the expressway industry overall.

According to a survey by Adobe Commerce (Adobe Communications Team, 2022), 67% of consumers prefer receiving personalized offers that align with their spending patterns, whether their shopping is done online or in physical stores. Personalization represents a significant opportunity to meet and even surpass customer expectations. To personalize services for each customer, the Expressway Organization established a dedicated operational team responsible for data analysis. Identifying customer needs is a primary objective for predictive efforts, yet these needs present a challenge due to their lack of clear definition and frequent changes.

Due to the unique nature of car segmentation data and the time-consuming process of analyzing data using machine learning, which involves many people, this research focuses on solving both problems by presenting a hybrid algorithm combined with AutoML to save time and produce more accurate models. The drawbacks of traditional machine learning processes include lengthy execution times, the need for human involvement at each stage, and a systematic approach to selecting the most suitable model for a given project. In summary, machine learning demands expertise and lacks automation. AutoML seeks to automate aspects of the machine learning model-building process; however, currently, its capabilities are primarily focused on classification and regression models.

An Automated Hybrid Machine Learning Framework

This research introduces the AHML framework as a novel solution for the expressway industry. This model incorporates hybrid algorithms grounded in machine learning, encompassing clustering and classification with an automated process. Utilizing the latest AHML framework concept, which applies AutoML, which is a part of artificial intelligence (AI), to facilitate the rapid and automated creation of machine learning models, this research applies this new concept to the car segmentation of Thailand's expressway industry to expedite model construction in a more automated manner.

Figure 1 shows the AHML architecture that adopts the AHML framework. The AutoML framework is the latest concept that leverages AI to automatically create models in machine learning. It addresses a pain point in the data analytics industry that typically necessitates experts creating models through various steps. With this framework, the time and effort required to create a model are significantly reduced.

Table 1 shows numerous AutoML products available on the market. Currently, the AutoML framework is proficient at classification and regression models; therefore, researchers are applying a clustering model developed by the marketing team. The contribution of the research is to integrate both clustering and classification into the AHML framework with automated ability.

For this study, the researchers utilize auto-sklearn due to its popularity and seamless integration with Python. Auto-sklearn is a widely adopted open-source package used by researchers worldwide for hypothesis testing and model-building purposes.

Hybrid Classification Model

The machine learning module consists of two sub-modules: clustering and classification. The classification module serves as a core component in machine learning, leveraging labeled datasets to predict future outcomes. However, there are instances where class labels may be absent or not clearly defined. Clustering algorithms that represent unsupervised learning concepts contrast with classification, which is a supervised learning concept. Recently, many researchers have proposed hybrid classification models due to their ability to address shortcomings in typical classification models. These models often struggle with accurate predictions, especially when dealing with simplistic data structures. Conversely, clustering models naturally form real groups from data. In this research, a hybrid classification approach begins with clustering, and a classification model follows. The dataset, which includes expressway car transportation data, naturally exhibits clusters formed automatically