IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 12 January 2023, accepted 15 February 2023, date of publication 22 February 2023, date of current version 1 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247448

APPLIED RESEARCH

On Hyperparameter Optimization of Machine Learning Methods Using a Bayesian Optimization Algorithm to Predict Work Travel Mode Choice

MAHDI AGHAABBASI[1], MUJAHID ALI[2], MICHAŁ JASIŃSKI[3,4], (Member, IEEE), ZBIGNIEW LEONOWICZ[3,4], (Senior Member, IEEE), AND TOMÁŠ NOVÁK[4]

[1]Transportation Institute, Chulalongkorn University, Bangkok 10330, Thailand
[2]Department of Transport Systems, Traffic Engineering and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology, 40-019 Katowice, Poland
[3]Department of Electrical Engineering Fundamentals, Faculty of Electrical Engineering, Wrocław University of Science and Technology, 50-370 Wrocław, Poland
[4]Department of Electrical Power Engineering, Faculty of Electrical Engineering and Computer Science, VSB—Technical University of Ostrava, 708-00 Ostrava, Czech Republic

Corresponding author: Mahdi Aghaabbasi (mahdi.a@chula.ac.th)

This work was supported by the Studentská grantová soutěž [Student Grant Competition (SGS)] Grant from VSB—the Technical University of Ostrava under Grant SP2023/005.

ABSTRACT Prediction of work Travel mode choice is one of the most important parts of travel demand forecasting. Planners can achieve sustainability goals by accurately forecasting how people will get to and from work. In the prediction of travel mode selection, machine learning methods are commonly employed. To fit a machine-learning model to various challenges, the hyperparameters must be tweaked. Choosing the optimal hyperparameter configuration for machine learning models has an immediate effect on the performance of the model. In this paper, optimizing the hyperparameters of common machine learning models, including support vector machines, k-nearest neighbor, single decision trees, ensemble decision trees, and Naive Bayes, is studied using the Bayesian Optimization algorithm. These models were developed and optimized using two datasets from the 2017 National Household Travel Survey. Using several criteria, including average accuracy (%), average area under the receiver operating characteristics, and a simple ranking system, the performance of the optimized models was investigated. The findings of this study show that the BO is an effective model for improving the performance of the k-nearest neighbor model more than other models. This research lays the groundwork for using optimized machine learning methods to mitigate the negative consequences of automobile use.

INDEX TERMS Bayesian optimization algorithm, hyperparameters, sustainable mode choice decision, work travel mode choice.

I. INTRODUCTION

A significant portion of people's everyday journeys are related to their jobs. Attempts are being made by transportation researchers and designers to identify work-related travel habits and develop ways to decrease the negative effects of motorized transportation on traffic, wellbeing, and the ecosystem. Work travel mode choice (WTMC) is an essential activity that relates to the process of selecting a mode of transportation for one's commute to work. An extensive body of research has shown that a range of variables impact the choice of travel modes to work. The studies explore various factors that influence people's choices of transport mode for their daily commute, such as household characteristics, work-related factors, built environment, and socio-psychological determinants [e.g., 1, 2-10]. Many of the studies focus on specific regions or countries, including Shanghai, Dublin, Washington, DC, Delhi, Hanoi, the Greater Toronto area, and Sichuan [e.g., 1, 4, 9, 11-13]. Some studies examine differences in mode choice among different age groups and

The associate editor coordinating the review of this manuscript and approving it for publication was Tamas Tettamanti.

socio-economic backgrounds, while a few investigate the impact of the COVID-19 pandemic on daily activities and health parameters [15, e.g., 14]. Overall, the studies provide insights into the complex and multi-faceted nature of travel behavior and the factors that shape individuals' mode choice.

Despite the variety of topics and regions explored, many of the studies share a common goal of informing policy and planning decisions that can promote sustainable and healthy travel behavior [e.g., 5, 11, 12, 16]. By identifying the factors that affect mode choice, policy makers and planners can design interventions and infrastructure that encourage more sustainable and healthy travel, such as active transport modes like cycling and walking. The studies also highlight the importance of considering the diverse needs and preferences of different groups of people when designing transportation systems and infrastructure [17, e.g., 2, 3, 7, 8]. Ultimately, the findings of these studies can inform efforts to create more efficient, sustainable, and equitable transportation systems in communities around the world.

A broad variety of factors and samples are included in the mode choice data. According to. Aghaabbasi, et al. [18], usually, these data sets suffer from missing values and incomplete data. More individuals use vehicles to go to work than any other mode of transportation since motorized transportation is so prevalent across the globe. This leads to an uneven choice of travel modes in the survey results.

A crucial aspect of transportation planning and forecasting is the precise modeling of mode selection for different purposes, especially work travel [19]. Many studies to date have tried to investigate the feasibility of various methods for accurately predicting the choice of work travel mode. A summary of these studies is provided in Appendix A and Section II. The multinomial logit (MNL) model introduced by.McFadden [20] is unquestionably the most used approach for travel mode choice forecasting. In the context of the MNL model, decision probabilities have a closed form and are straightforward to comprehend [21]. But rigorous statistical requirements lead to the belief that a linear combination of variables captures all causes of variation and association across options [21], [22]. Non-linearity in the qualities of the options and the inclusion of causes of variability are both complex and require specialized knowledge in the MNL model construction [23], [24].

Compared to conventional statistical models, methodologies from the domain of Machine Learning (ML) are a viable option for modeling travel mode selection and provide an alternate method to conventional forecast modeling that may overcome present constraints. ML approaches describe complicated connections between variables by data-driven learning, as opposed to establishing strong assertions beforehand [25]. Numerous studies in the transportation domain have shown that ML approaches are more accurate predictors than traditional statistical techniques [e.g., 19, 26, 27-29].

In the majority of research that used ML algorithms to predict travel mode selection, the calibration of the ML models was not the result of a systematic search [30]. Developing a potent ML model needs time and effort since it entails locating the optimal algorithm and obtaining the optimal model architecture through fine-tuning hyperparameters [31]. ML models' parameters are divided into two categories: model parameters, which may be initiated and adjusted via data learning; and hyperparameters, which cannot be easily approximated from data learning and should be established prior to training an ML model since they determine the model architecture [32], [33], [34], [35]. When minimizing a loss function, hyperparameters are the parameters adjusted to optimize the model's performance [36]. Manual testing is the conventional method for tuning hyperparameters, and it is still widely utilized in a variety of fields of study, despite requiring a comprehensive grasp of the ML techniques and their hyperparameter value configurations. A high number of hyperparameters, complicated models, time-consuming model assessments, and non-linear hyperparameter relationships render manual tuning impractical for a variety of issues [37]. These considerations have stimulated a rise in research into strategies for the automated optimization of hyperparameters. The primary objective of hyperparameter optimization is to automate the hyperparameter tweaking process and enable users to efficiently apply ML models to real-world issues [31]. After a hyperparameter optimization procedure, it is anticipated that the ideal model structure of an ML model would be determined.

Several well-known methods are available for the hyperparameter optimization of ML models, including traditional methods (e.g., gradient descent-based methods), decision-theoretic methods (e.g., grid search and random search), metaheuristic methods (e.g., particle swarm optimization (PSO)), and Bayesian methods. Most hyperparameter optimization problems are non-convex or non-differentiable optimization problems and might even lead to a local rather than a global optimum, which renders traditional optimization approaches inappropriate for hyperparameter optimization problems [38]. These methods, like gradient descent-based methods, can be used to optimize continuous hyperparameters by figuring out their gradients [39]. For hyperparameter optimization, decision-theoretic approaches, metaheuristic algorithms, and Bayesian optimization (BO) models outperform traditional optimization methods such as gradient descent [40]. Many of these methods can find discrete, categorical, and conditional hyperparameters just as well as continuous ones. Time is a major factor in decision-theoretic methods due to their blind nature. Concerning the metaheuristic approaches such as PSO, it is simple for these algorithms to reach local optimums in high-dimensional space, and they have a low convergence rate during the iterative procedure [41], [42]. BO models estimate every next hyperparameter value based on the outcomes of previously tested hyperparameter values, hence reducing the number of superfluous assessments. Therefore, BO is able to identify the ideal hyperparameter configuration with fewer repetitions than grid search and random search [43]. BO can be used for a wide range of problems because it can use different surrogate functions, like the Gaussian process (GP), to describe how the objective function is distributed [44], [45], [46], [47].

The purpose of this paper is to assess how the BO algorithm can improve the performance of ML models to predict work travel mode choice. To this end, the performance of five ML models, including support vector machine (SVM), k-nearest neighbors (KNN), single decision trees (SDT), ensemble decision trees (EDT), and Naive Bayes (NB), was assessed while their hyperparameters were optimized using the BO algorithm. These models were developed and optimized using two datasets from the US 2017 National Household Travel Survey (NHTS) and several partition portions. The performance of these models was assessed using average accuracy and average area under the receiver operating characteristics curve (AUC).

The format of the following sections of this work is as follows: The second section offers an overview of the research and planning that preceded the data collection. The study's methodology is described in the third section. The fourth section provides an overview of the ML models whose hyperparameters were optimized in this study. The fifth section describes the input selection. The sixth section presents the models' development and evaluation. The seventh part offers a concise summary of the research.

II. LITERATURE REVIEW

The studies mentioned in Appendix A present a diverse set of factors that influence work travel mode selection. Al-Ahmadi [48] identified travel time, cost, household income, and car ownership as essential factors in the intercity work mode choice model for Saudi Arabia. Badoe [3] claimed that household-level choice models better predict mode choice for two-worker households and that multinomial logit is better suited for testing the approach. In addition, the study showed that household models better predicted mode choice for two-worker households than individual models. Day et al. [4] showed that analyzing trip timing and mode choice together in travel demand models, by developing multinomial logit models for different occupation groups, can reveal significant differences in mode choice preferences. Gang [1] estimated the utility function and calculated time and choice probability elasticities for Shanghai work-trip mode choice behavior using multinomial choice models. The research found that bus, subway, and taxi users prioritize "in-vehicle time," "out-of-vehicle time," and "money cost," and that bikes are superior for all levels of income. Habib [17] developed a joint trivariate model for commuters' mode choice, work start time, and work duration, which is estimated using data from the Greater Toronto Area. The model is intended to capture relationships among random components impacting these choices and is used to forecast employees' work schedules according to mode choice, hence offering insights into the behavioral intricacies of mode choice and work scheduling. Hamre and Buehler [12] found that free car parking at work is related to more driving, while commuters offered benefits for public transportation, walking, or cycling but no free car parking are more likely to use those modes. In addition, the study found that benefits for public transportation, walking, and cycling seem to work best when car parking is not free.

Hatamzadeh et al. [49] developed behavioral choice models to explore the factors influencing walking behavior for work and shopping trips in Rasht, Iran. The study found that different factors influenced the propensity to walk for males and females, with the presence of young children in households, time of day, and distance influencing the propensity to walk. Heinen and Bohte [50] investigated the factors influencing the combined use of public transportation and bicycles for commuting, identifying attitudes toward mode choice as significant in the decision to commute by both public transportation and bicycle. In addition, the study found significant differences in beliefs about public transport and cycling between the groups. Heinen et al. [5] found that a favorable attitude toward cycling and coworkers' expectations were connected with a greater probability of becoming a commuter biker, but the existence of facilities for alternative transport modes and an increase in trip length lowered the likelihood of riding. Indriany, et al. [51] showed that travel time attributes are essential for understanding the influence of uncertainty in networks on commuter mode choice in the South Tangerang and Jakarta regions. Irfan et al. [52] developed a travel behavioral model for work-trip mode in Rawalpindi, Pakistan, using the multinomial logit model. The study found that travel demand is elastic with respect to congestion pricing and out-of-vehicle travel time. Furthermore, the study showed that congestion pricing is an effective means of reducing automobile demand, and when combined with improved transit services through BRT, it can induce a greater modal split than either policy alone. Kunhikrishnan and Srinivasan [53] revealed that Chennai's working population's mode choices are heterogeneous due to variances in option sets, natural preferences, and susceptibility to explanatory variables.

The studies use a range of modeling and analytical methods to investigate work travel mode selection. These include multinomial logit models, binary logit models, binomial logit models, spatial general equilibrium models, and joint trivariate models. A few studies used advanced ML methods. However, the risks associated with not using advanced ML methods depend on the problem being addressed. If the problem is relatively simple and can be solved with traditional statistical methods or heuristics, then not using ML methods may not be a significant risk. However, in many real-world applications (e.g., mode choice data), the data is large, complex, and high-dimensional, making it difficult to manually create effective models or heuristics. In such cases, not using advanced ML methods can result in inaccurate predictions or decisions, leading to missed opportunities or even potential harm. Additionally, not using advanced ML methods can lead to suboptimal solutions or inefficient processes, which can be a disadvantage in highly competitive markets.

III. DATA

The SVM, KNN, SDT, EDT, and NB models and their optimized variants developed in this study were applied into two datasets. These datasets belonged to the US 2017 National Household Travel Survey (NHTS). The Federal Highway Administration conducts the NHTS, which is the official source on American public travel patterns. Individual and household travel may only be studied using this national large dataset. Everyday non-commercial travel by any method is included in NHTS. This includes information on travellers, their families, and the cars they use.

As previously mentioned, the NHTS covers all US states. Two US states have been randomly selected for testing the SVM models in this study. These datasets include Iowa and Ohio. Randomly picking states guarantees that ML models are tested on a sample of data that is representative of the whole population. This is due to the fact that diverse state features, such as population density, demography, and economic considerations, might affect the model's performance. By picking states at random, it is more probable that this study's sample will reflect the complete range of data features. In addition, random sampling prevents researchers from picking states that may be more favorable to their model, which might lead to overfitting. If researchers choose states based on their existing knowledge or opinions, this may introduce bias into the assessment process and make it difficult to apply the conclusions to new data.

Each of the selected datasets included many variables. Following a comprehensive examination of the literature, the researchers in this study focused on characteristics connected to the mode of transportation used to get to work. Ultimately, there were 22 variables in all, including 21 input variables and one target variable (travel mode choice to work). Table 1 contains a list of these variables, as well as brief descriptions of each. A statistical analysis of the variables used in this study can be found in Appendix B. The datasets, the number of records they have, and the imbalance ratio are all shown in Table 2.

TABLE 1. List of variables.

Variable	Description	Type
AGE	The respondent's age	CO
EDUC	Level of education	CA
GT1JBLWK	Several jobs	FL
FLEXTIME	Work start time flexibility	FL
RACE	Race	NO
SEX	Gender	FL
WKFTPT	Working full-time or part-time	FL
HOMEOWN	Ownership of a home	FL
DRVRCNT	Household's number of drivers	CO
HH_ONTD	Number of household members on trip	CO
HHFAMINC	Income of household	CA
HHSIZE	The total number of people living in the home	CO
HBPPOPDN	Population density in the census block category in which the household resides	CA
HHVEHCNT	The total number of automobiles owned by a household.	CO
NUMADLT	Total number of adults in the household who are at least 18 years old	CO
VEHOWNED	Possessed a car for more than a year	FL
WRKCOUNT	The number of employees in a household	CO
YOUNGCHILD	Number of children aged 0 to 4 living in the household	CO
TIMETOWK	Time spent travelling to work	CO
URBANSIZE	The size of the urban area in which the residence is situated	CA
URBRUR	A household in an urban or rural setting	CA
WRKTRANS*	Travel Mode to work	FL

*Target variable

Continuous: CO; Flag: FL; Categorial: CA; Nominal: NO

TABLE 2. States' number of records, record distribution, and imbalance ratios.

State	Records	Distribution of records (%)		Imbalance ratio
		Motorized (1)[1]	Non-motorized (2)[2]	
Ohio	560	547 (97.68%)	13 (2.32%)	42.08
Iowa	1515	1473 (97.23%)	42 (2.77%)	35.07

[1]car; [2]walking and biking

IV. DESIGN OF STUDY

This research combines the BO algorithm with five established ML methods—SVM, KNN, SDT, EDT, and NB—to predict the mode of transportation used to and from work. As was indicated before, the information used to create this research came from the 2017 NHTS in the United States. The Gaussian Process (GP) model was initialized with 30 evenly produced random seed points. This method ensures that the model examines a varied range of hyperparameters early in the optimization process, which might result in a quicker convergence to the optimum set of hyperparameters. By choosing equally spaced points, the model may avoid being trapped in local optima and offer a more accurate initial approximation of the function being optimized. This may aid in accelerating the optimization process and providing superior outcomes. Minimizing the 5-fold cross-validation loss is the objective. The acquisition function is "expected-improvement per second plus" in order to avoid local minima. 100 is the maximum number of assessments allowed. The datasets are separated into three distinct portions with corresponding training and test ratios of 70:30, 80:20, and 90:10. The samples of these partitions were randomly selected. Average accuracy (%) and average area under the ROC curve (AUC) were used to evaluate the efficiency of the models constructed in this investigation. A schematic representation of this investigation is shown in Figure 1.

A. BAYESIAN OPTIMIZATION PRINCIPLES

BO iteratively builds a probabilistic model of the function being optimized and selects the next set of hyperparameters to assess according to the current best estimate of the function's behavior to obtain the optimum set of hyperparameters for an ML model. The fundamental tenets of BO may be summed up as follows: (1) define a prior distribution; (2) select the next set of hyperparameters to evaluate; (3) evaluate the function; (4) update the probabilistic model; and (5) repeat steps 2-4. By repeatedly updating the probabilistic model, BO quickly

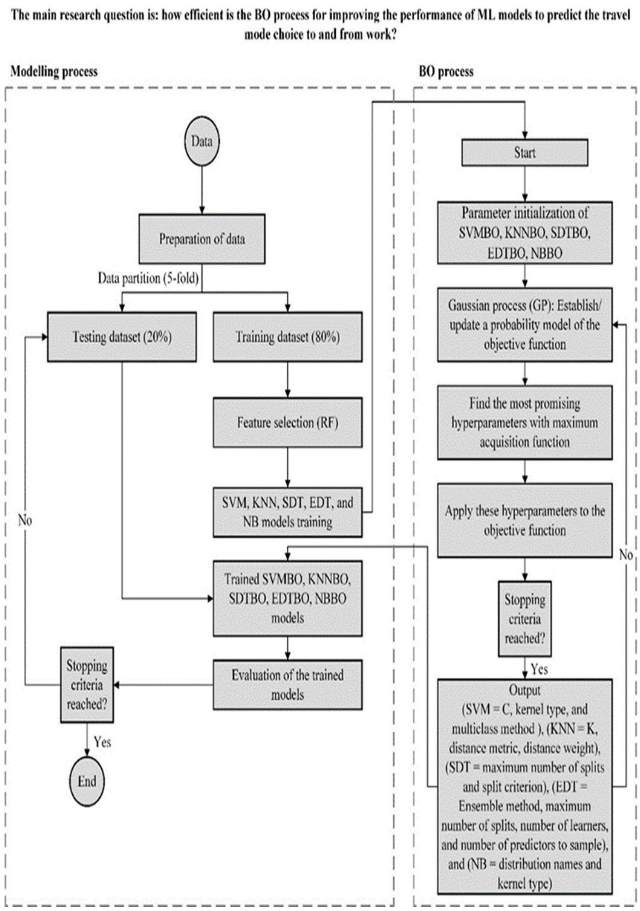


FIGURE 1. Flowchart of this study.

examines the space of hyperparameters and converges to the optimum set with a minimum number of function evaluations.

B. PERFORMANCE EVALUATION CRITERIA

The effectiveness of the models developed in this research was evaluated using two criteria. These criteria include accuracy (%) and AUC. Prediction accuracy is expressed as the number of accurate predictions across two classes divided by total number of predictions. As the name implies, the AUC metric measures the classifier performance by measuring the area under a receiver opera-tor characteristic curve. For testing and training datasets, the accuracy and AUC are averaged over two datasets and three partition portions.

V. THE BACKGROUND OF THE ML MODELS EMPLOYED IN THIS STUDY

A. SUPPORT VECTOR MACHINES (SVM)

By segmenting the information into linear and nonlinear structures, SVM is able to generate a reliable decision boundary [54]. The SVM separates the information into different categories depending on the margin with the highest value, which represents the distance between the border and the data (Figure 2). When compared to other methods of ML, SVM is more efficient and straightforward. The hyperparameters

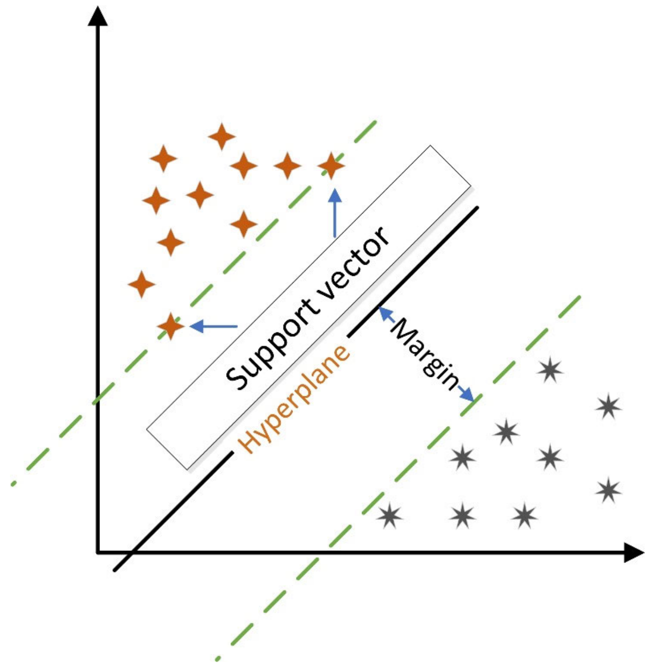


FIGURE 2. Graphical representation of the SVM model.

may be adjusted to provide the optimum boundary conditions, and developers can choose kernel functions intuitively. This hyperparameter, which relies on the distribution of the encompassing data, determines the margin and kernel size that make up the border's overall form. Three hyperparameters (the kernel function, the box constraint level, and the multiclass technique) were subjected to BO in order to improve the SVM forecasting model's efficacy.

B. K-NEAREST NEIGHBORS (KNN)

The k-Nearest Neighbors (KNN) technique is a popular choice since it uses a similarity measure to assign classes to newly acquired records (Figure 3) [55]. In practice, it is often used to assign a label to a data point depending on how its neighbors are labelled. In KNN, the parameter K specifies the maximum number of nearest neighbors to use for determining a majority. K is calculated by how well each object's characteristics match up with one another. Parameter tuning refers to the process of determining the optimal value of K in order to enhance performance. Using lower values of K increases the likelihood of making errors that will have a bigger effect on the final result. In addition to lower variance and greater bias, larger K values are also related to smoother decision boundaries. Timing is also an issue with this. Not only must K, but also the distance metric and distance weight, be optimized while using KNN.

C. SINGLE AND ENSEMBLE DECISION TREES

A single decision tree (SDT) is a structured-tree classifier that iteratively applies decision nodes and leaf nodes to fulfil classification tasks. This method is easy to understand, easy to implement, can handle missing values well, and learns

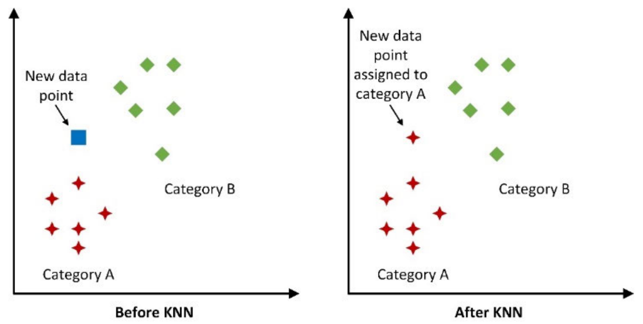


FIGURE 3. The Schematic system of the KNN model.

quickly. However, SDT is susceptible to overfitting and has poor nonlinear data processing flexibility. The result is very sensitive to even minute changes in the input, and the DT subtrees may be duplicated many times. SDTs have two important hyperparameters that should be optimized, including the maximum number of splits and the split criterion, which will be optimized using BO.

The ensemble decision tree (EDT) methodology is a ML method that employs several SDTs constructed from the provided data. EDT may be assembled using a variety of methods, including bagging and boosting (Figure 4). Bagging randomly chooses data from a training dataset to make many sets. Each set is then used to train its own DTs. By fitting basic models to the data, early learners are utilized to train a number of weak learners sequentially in the boosting technique. The weights are raised to account for any inaccuracies in subsequent analyses. When com-pared to an SDT, employing the average and forecasted values of several groups of trained trees significantly lowers overfitting and decreases prediction uncertainty. The final projected value in tree ensemble regression is obtained by either averaging the predictions of all DTs (called "bagging") or by using the predictions of a strong learner constructed from several weak learners (called "boosting"). In the EDT-based work travel mode choice pre-diction model, BO is used to look at the ensemble technique, the maximum number of splits, the number of learners, and the number of predictors to sample.

D. NAIVE BAYES (NB)

NB [56] is a straightforward ML approach that uses the Bayes theorem to determine class probabilities under the assumption that the characteristics are unrelated to one another. Then, predictions for the class with the greatest likelihood are produced. To compute probabilities from continuous features, the probability distributions of those characteristics must be approximated. Commonly, kernel density estimation is used for this purpose. Despite the fact that the independence assumption of NB seldom holds true in reality, the classifier has been found to be comparable with more powerful classifiers. The NB hyperparameters that need to be optimized using the BO algorithm are distribution names and kernel type.

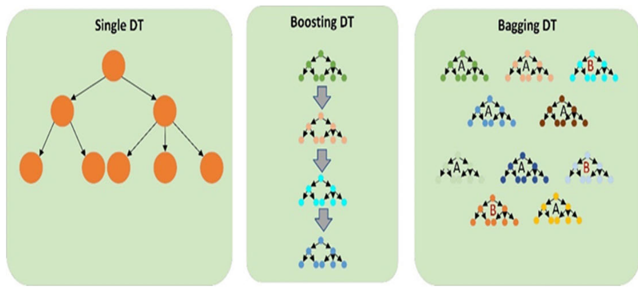


FIGURE 4. Schematic diagram of SDT and EDT.

VI. FEATURE SELECTION USING RANDOM FOREST (RF)

Prior to building and optimizing the ML models, the researchers used the random forest (RF) [57] approach to pick the most relevant features. The RF algorithm and its ensemble theory are widely used to analyse large amounts of data, select features, and predict the variables of interest. Weak-regressors are created by the RF method using decision trees with inputs that are either randomly selected from a training dataset or sampled from a random subset. Data segmentation and decision-making models are learned in each and every DT via a given dataset. Instead, characteristics in their dataset are examined and disassembled in order to arrive at a sound conclusion. All the model's forecasted decision outputs are fed into the regression algorithm throughout this procedure. Final forecasts are determined using the average of all estimates.

RF can estimate the relevance of inputs in a regression or classification problem in a straightforward way. In the first stage, the data set $K_n = \{(A_i, B_i)\}_{i=1}^{n}$ is fitted into an RF model. "A" represents a training set, whereas "B" represents the matching response set. The out-of-bag (OOB) error for each data point is recorded and then averaged throughout the whole dataset before the fitting method continues. After training is complete, the OOB error is assessed several times on a dataset with the Nth input's values swapped around to calculate the Nth input's order. By averaging the differences in OOB error before and after the conversion throughout all trees, an importance score for the Nth input may be determined. The last step in the RF's normalization process uses the standard deviation of these differences to establish a score. Greater significance is given to inputs that result in higher values for this score than to those that result in lower values for this score. Figure 5 depicts a basic RF procedure for deriving predictions and significance ratings.

RF was applied to two datasets, including Iowa and Ohio to select the most relevant inputs for predicting work travel mode choice in each dataset. In the Iowa dataset, six inputs were picked, and seven inputs were selected in Ohio. Figure 6 illustrates the selected inputs. Five ML models and their optimized variants for each US state will be developed using these inputs. The researchers used a variety of factors and settings to execute this experiment (Figure 6).

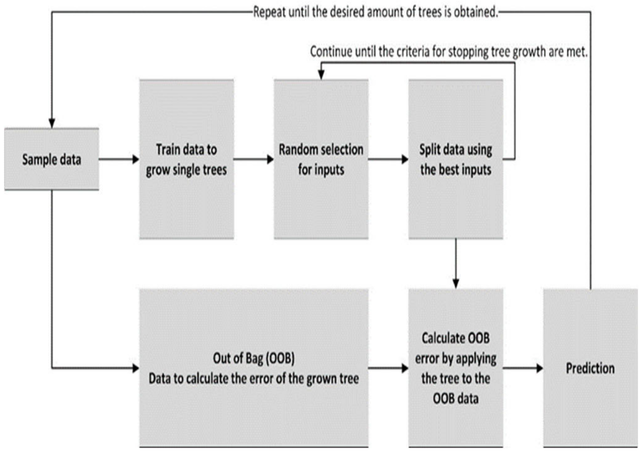


FIGURE 5. RF schematic diagram.

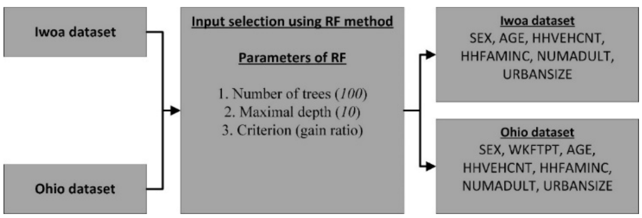


FIGURE 6. Selected inputs using RF and its parameters.

VII. RESULTS AND DISCUSSIONS

A. DEVELOPMENT AND EVALUATION OF THE OPTIMIZED MODELS TO PREDICT WORK TRAVEL MODE CHOICE

Using MATLAB, the BO algorithm is used to optimize the hyperparameters of SVM, KNN, NB, EDT, and SDT for two datasets from the 2017 NHTS in the United States. The GP model was initialized with 30 uniformly generated random seed points. Minimizing the 5-fold cross-validation loss is the objective. To avoid the traps of the local minima, the acquisition function is "expected-improvement per second plus." The models underwent 100 rounds of iteration. For training and testing ratios, the datasets were split into three parts: 70:30, 80:20, and 90:10. In Figures 7 and 8, the minimal classification error vs. the number of iterations is shown on a graph. All optimized models in two datasets (with different split portions) are demonstrated to need fewer than 100 iterations to reach their minima. For this reason, developing SVM, KNN, NB, EDT, and SDT using a variety of data types is greatly facilitated by the BO method. It should be noted that the optimized values of the ML models' hyperparameters are shown in Appendix C and D. The average number of iterations needed to train the optimal ML models by different partitioning was also compared. Figure 9 illustrates the outcomes of this comparison. For 70:30 partitioning, KNNBO and SVMBO were the quickest models; for 80:20 partitioning, SVMBO and NBBO had the quickest convergence; and for 90:10 partitioning, SVMBO was the quickest model to obtain the least classification error. Overall, SVMBO was the quickest model to train across all partitions for predicting the mode of work-related travel.

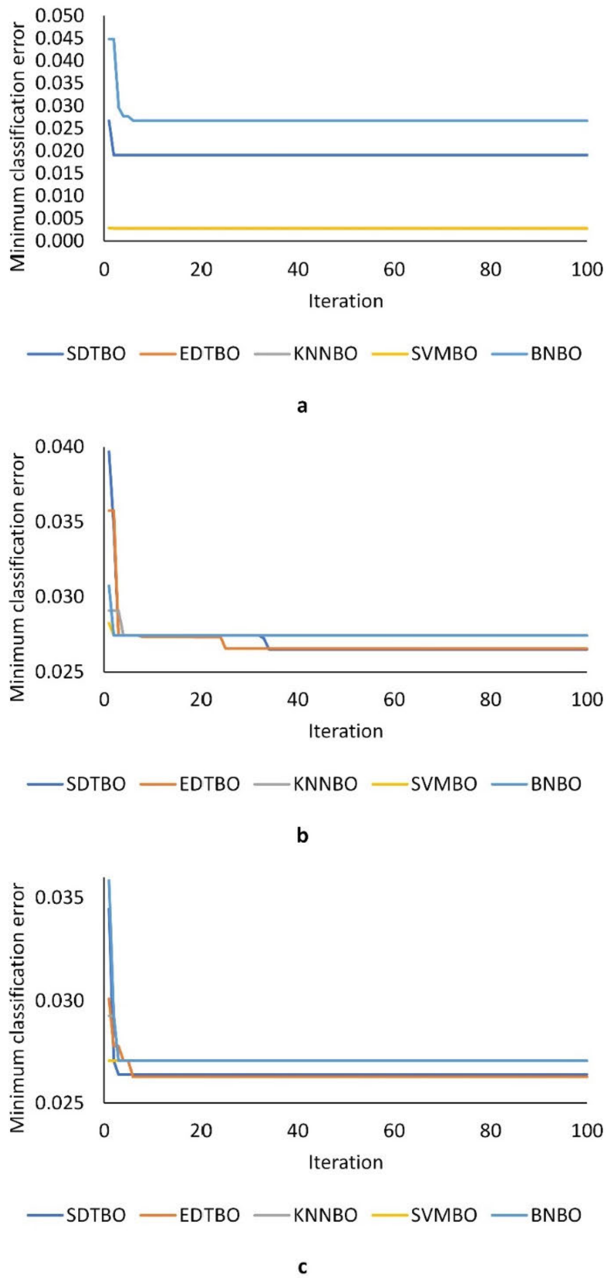


FIGURE 7. Iteration process of ML models' optimization-Iowa dataset: (a) 70:30 training/testing partitioning; (b) 80:20 training/testing partitioning; (c) 90:10 training/testing partitioning.

We calculated the optimized models' average accuracy and AUC across all partitions and both datasets. The outcomes of this calculation are shown in Table 3. We also used a simple ranking system to rank the average accuracy and AUC in each of the training and testing phases. These ranks were then summed up to achieve the total ranking. As can be seen, KNNBO had the best average accuracy and AUC during training, as well as the best average AUC during testing.

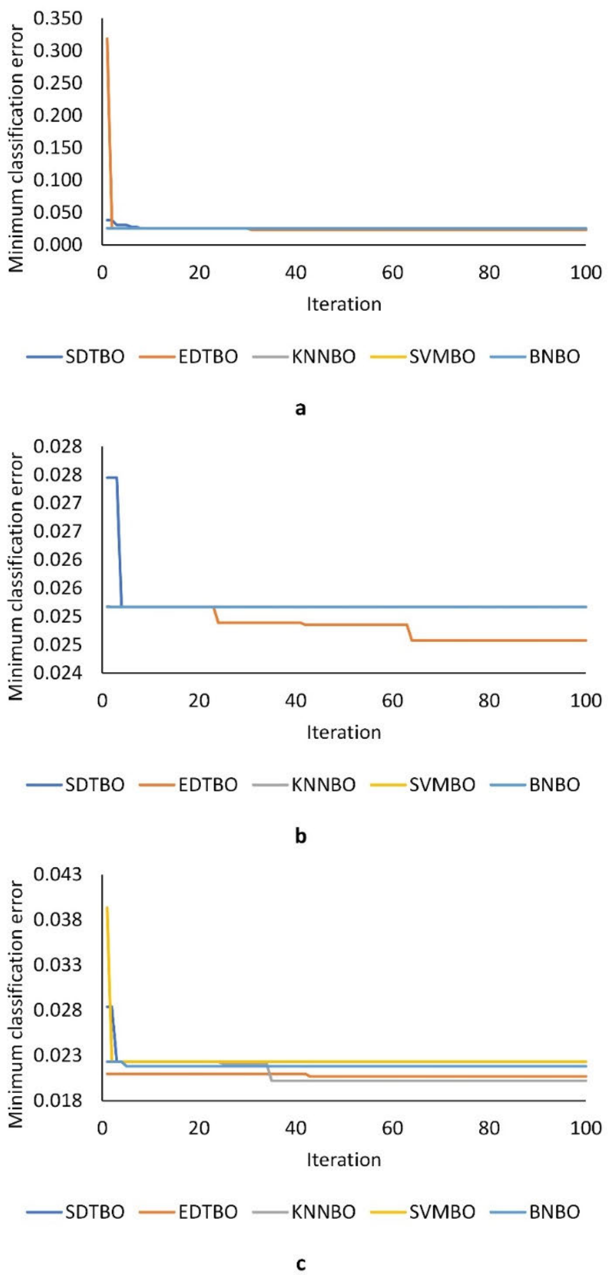


FIGURE 8. **Iteration process of ML models' optimization-Ohio dataset: (a) 70:30 training/testing partitioning; (b) 80:20 training/testing partitioning; (c) 90:10 training/testing partitioning.**

Consequently, the KNNBO model achieved the highest total ranking in terms of accuracy and AUC for predicting the work travel mode choice.

Figure 10 shows the average changes in accuracy and AUC for two datasets. Figure 10.a demonstrates that the application of BO to the KNN model produced the best average gain in accuracy throughout both the training (+1.683 percent) and testing (+1.067 percent) phases. Figure 10.b further demonstrates that the KNN model optimized by the BO method had the greatest increase in both training and testing AUC (+0.103 and +0.192, respectively).

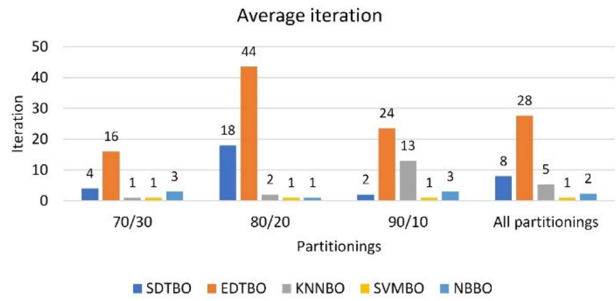


FIGURE 9. **Changes in average accuracy (%) and AUC across two datasets: (a) changes in accuracy; (b) changes in AUC.**

This study also examined the impact of the BO algorithm on the average accuracy improvement of the prediction of the non-motorized class, which is the minority class in this study. To this end, the average testing prediction accuracy of SVM, KNN, EDT, SDT, and NB before and after optimization and across the two datasets and the three data partitionings in each dataset were calculated and compared. Table 4 shows the results of this comparison. It can be seen that the hybridization of the SVM and BO, as well as the hybridization of the EDT and BO, yielded the best average prediction accuracy for the minority class. While the KNNBO obtained the best ranking among various ML models developed in this study, the EDT and SVM models outperformed the KNN model in terms of improving the prediction accuracy of the minority class when they were hybridized with the BO model. One possible reason for this is that KNN is a relatively simple algorithm that relies on the distance between data points and may not be the best algorithm for complex or poorly separated data, especially for the minority class. Another possible reason is that the hyperparameters of KNN are not the most critical factors for improving the prediction accuracy of the minority class, and thus BO did not significantly improve the accuracy of KNN with regards to the minority class.

We also compared the results of models' hyperparameter optimization using the BO algorithm with the results of optimizing using a grid search strategy. Figure 11 shows the average testing accuracy that was reached across the two datasets and the three partitions in each dataset when BO and grid search optimization methods were used. The findings show that BO is more effective than grid search for optimizing ML models and improving the accuracy of the models for predicting the work travel mode choice.

The findings of this comparative study show that BO was more efficient at improving the performance of the KNN models than other ML models to predict work travel mode choice. However, the KNNBO model was not successful in improving the overall accuracy of the minority class. In addition, the KNNBO model was not the fastest to reach the minimum classification error. However, the average number of iterations it took to get there was only five, which is a low number. For ML models with a small-sized continuous

TABLE 3. Optimized models' average accuracy, AUC, and total ranking.

Model	Ave. train accuracy	VR	Ave. test accuracy	VR	Ave. train AUC	VR	Ave. test AUC	VR	TR
NBBO	97.43	2	97.38	5	0.70	4	0.63	3	14
SVMBO	97.83	4	97.38	5	0.60	2	0.54	2	13
KNNBO	97.87	5	97.23	3	0.71	5	0.69	5	18
SDTBO	97.58	3	97.17	2	0.59	1	0.54	2	8
EDTBO	97.83	4	97.28	4	0.69	3	0.67	4	15

Value ranking = VR; total ranking = TR

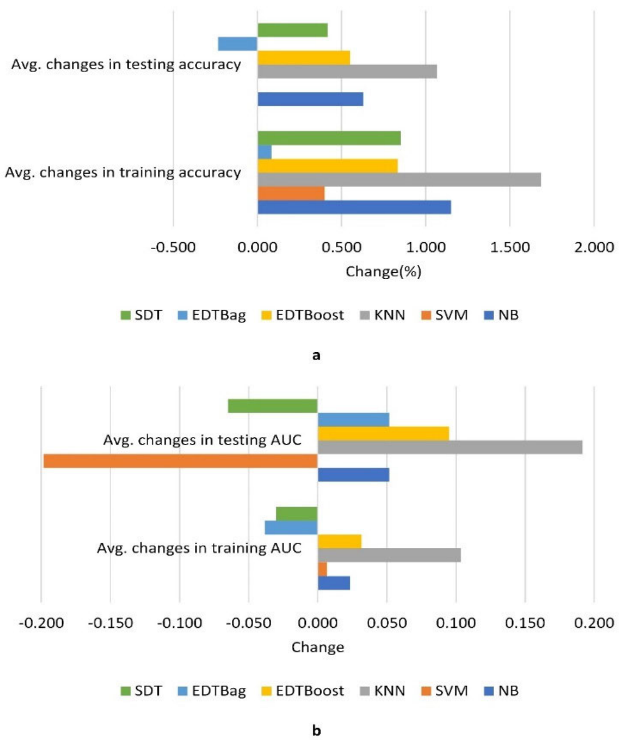


FIGURE 10. A comparison between the average number of iterations required for training the optimized models.

hyper-parameter space, like KNN, BO is often able to gain improved performance metrics [37]. It also should be noted that all models optimized by BO achieved high training and testing accuracy, which shows the capability of this algorithm to yield models with high accuracy.

The BO algorithm performs well because it uses information from previous iterations to help choose the optimal parameters for the next one. That is probably the main reason why the BO can help the ML algorithms make desirable predictions for the work travel mode choice. Using the data gathered from prior evaluations, BO constructs a probabilistic model that maps hyperparameters to objective function scores. With this method, the surrogate probability model

TABLE 4. Impact of BO on the prediction accuracy of the minority class.

	Before optimization	After optimization	Changes (%)
	Average accuracy (%)	Average accuracy (%)	
SVM	5.56	16.66	11.1
KNN	22.22	8.3	-13.92
EDTBag	28.97	22.22	-6.75
EDTBoost	0	22.22	22.22
NBBO	27.64	8.33	-19.31
STDBO	33.2	8.33	-24.87

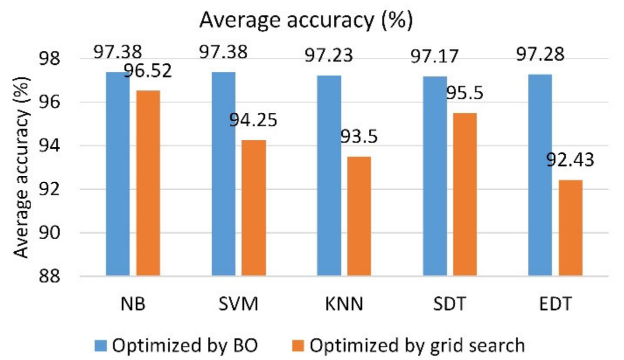


FIGURE 11. The average accuracy of various ML models developed in this study was optimized by BO and grid search.

is continuously updated after each evaluation of the target function, making it "less inaccurate" as more data is accumulated. Since BO uses this information to fine-tune the hyperparameters that come after, it is indeed effective. One key idea underlying BO is to spend more time choosing the remaining hyperparameters in order to cut down on the total number of times the objective function is invoked. In actuality, the effort spent on selecting the following hyperparameters is insignificant in contrast to the time spent on the objective function.

This study mostly got better results than previous studies that used standard forms of ML models to predict work travel mode choice. All optimized models in this study achieved higher training and testing accuracy than those of Xie et al. [58] Qian et al. [34], and Hagenauer and Helbich [19], which used standard forms of various ML models. Concerning recent studies that employed optimized models to predict general mode choice, the present study achieved better results than those of. Kashifi et al. [59] who used EDT and SDT.

Considering implications for policymaking, optimized ML approaches like KNN, SVM, EDT, SDT, and NB and their optimized variants may be used by transportation researchers to find hidden patterns in urban journeys and travel behavior, as well as create predictions. These insights and forecasts may aid decision-makers in devising optimum ways to improve the reliability and efficiency of transportation networks. Optimized ML algorithms can read people's trip data, find out how they usually travel, and predict how they will travel in the future. Despite the significant benefits of ML approaches, researchers and policymakers have to overcome a wide range of problems, including data availability. This research shows that the BO algorithm with the ML model has various characteristics that make it beneficial to transportation policy and decision-makers. The suggested models are adequately resistant to changes in the datasets. As a consequence, decision-makers may readily tweak any input variable and see how the outcomes change. It is also possible to extend the model and include new information by simply revising the frequency tables for every input factor.

VIII. CONCLUSION

Transport and urban planners may be able to design more sustainable urban transport systems if they can accurately forecast travel mode choice to work. In this study, we built and optimized five ML-based models (SVM, KNN, EDT, SDT, NB) to predict the work travel mode choice by settling on the optimal set of hyperparameters for each ML technique. The accuracy of the models' predictions was investigated. This research made use of data from the 2017 NHTS in the United States. Models were tuned to boost performance consistently, and statistical measurements were employed to establish how well they predicted outcomes. Before developing these models, this study selected the most relevant inputs using RF. The models and their optimized variations were created using six input variables from the Iowa dataset and seven input variables from the Ohio dataset. We may summarize the results as follows:

- The SVM model made minimum classification errors on average faster than other models did over repeated iterations.
- The KNNBO model achieved the highest ranking in terms of average accuracy and AUC.
- In the EDT and SVM models, the BO method was effective in enhancing the prediction accuracy of the minority class.
- The BO approach enhanced the classification accuracy and AUC of both the training and testing stages for the standard KNN model.

This paper's results demonstrate that the BO approach can boost the capability of baseline ML models to forecast work travel mode choice. A higher improvement was achieved by including BO into the KNN model. Future research is encouraged to prioritize this model above other models when attempting to forecast the work travel mode choice.

In order to better train the ML algorithms used to predict the work travel mode choice, this work made use of two datasets, as was previously noted. The authors are certain that this sample size of datasets will allow them to accomplish their research goal. Nonetheless, additional datasets might be used in the future to examine the BO algorithm's potential in enhancing the performance of other ML algorithms in forecasting travel mode choice. It is possible that the BO method will be used in the future to find solutions to additional transportation problems, such as determining the optimal trip time, cost, or mode of transportation to get people from one place to another. Future studies can employ more advanced methods like deep learning and metaheuristics to analyse the mode choice data and compare their findings with those of this present study. This research employed RF method for input selection in this field. This method may be used by other researchers to identify the most important contributors to their dependent variables and facilitate the creation of simpler models.

APPENDIX A
A SUMMARY OF THE STUDIES ON THE TRAVEL MODE CHOICE TO WORK

Reference	Main Factors Used	Modelling Method
[48]	CR, SE, and SS	Disaggregate models and utility maximization
[3]	SD	MNL
[4]	WTC, WA, EP, and TDM	MNL
[60]	TDM	Spatial general equilibrium
[1]	SE	MNL
[17]	WA	MNL
[12]	TDM	MNL
[49]	SD	BL
[50]	AT	MNL
[5]	WA	BL
[51]	SS	BNL
[52]	Econometric modeling	MNL
[53]	Contextual discrepancy	BL
[61]	LOS	Class Association Rules
[62]	SD	MNL
[34]	SD, WTC, and BE	SVM
[2]	SD, SE, and BE	ZINB
[6]	BE	MNL
[63]	EP	GIS and CNL
[11]	WTC and BE	BL and GIS
[58]	SD and LOS	MNL, DT, and NN

Models - multinomial logit = MNL; binary logit = BL; binomial logit = BNL; decision trees = DT; neural networks = NN; cross-nested logit = CNL; zero-inflated negative binomial regression = ZINB; support vector machines = SVM
Factors - work attributes: WA; sociodemographic: SD; socioeconomic: SE; level of service: LOS; built environment: BE; work travel characteristics: WTC; travel demand management: TDM; employment patterns: EP; attitudes: AT; safety and security: SS; cultural and religious: CR

APPENDIX B
THE STATISTICAL ANALYSIS OF THE VARIABLES USED IN THE THREE DATASETS

Variable	Type	Ohio				Iowa			
		Min	Max	Mean/mode	Std. Dev	Min	Max	Mean/mode	Std. Dev
AGE	CO	16	86	45.22	14.97	16	87	43.92	15.26
EDUC	CA	1	5	4	-	1	5	4	-
GT1JBLWK	FL	1	2	2	-	1	2	2	-
FLEXTIME	FL	1	2	2	-	1	2	2	-
RACE	NO	1	6	1	-	1	6	1	-
SEX	FL	1	2	2	-	1	2	2	-
WKFTPT	FL	1	2	1	-	1	2	1	-
HOMEOWN	FL	1	2	1	-	1	2	1	-
DRVRCNT	CO	1	5	2.038	0.77	1	5	1.99	0.74
HH_ONTD	CO	1	6	1.39	0.718	1	6	1.41	0.78
HHFAMINC	CA	1	11	6	-	1	11	6	-
HHSIZE	CO	1	7	2.60	1.21	1	8	2.50	1.17
HBPPOPDN	CA	50	17000	3000	-	50	30000	7000	-
HHVEHCNT	CO	1	7	2.39	1.19	1	10	2.28	1.14
NUMADLT	CO	1	5	2.04	0.74	1	5	1.95	0.68
VEHOWNED	FL	1	2	1	-	1	2	1	0
WRKCOUNT	CO	1	4	1.793	0.74	1	5	1.78	0.71
YOUNGCHILD	CO	0	2	0.15	0.45	0	3	0.16	0.46
TIMETOWK	CO	0	290	23.84	20.27	0	200	17.70	15.24
URBANSIZE	CA	1	6	4	-	1	6	2	-
URBRUR	FL	1	2	1	-	1	2	1	-
WRKTRANS	FL	1	2	1	-	1	2	1	-

Continuous: CO; Flag: FL; Categorial: CA; Nominal: NO
Mean is used for continues variables; mode is used for nominal, flag, and categorial variables

APPENDIX C
ML MODELS' OPTIMIZED HYPERPARAMETERS FOR THE OHIO DATASET

Model	Hyperparameter	70:30	80:20	90:10
NB	Distribution names	kernel	kernel	kernel
	Kernel type	box	epanechnikov	box
KNN	Number of neighbours	1	216	1
	Distance metric	jaccard	mahalanobis	correlation
	Distance weight	equal	equal	inverse
SVM	Kernel function	quadratic	gaussian	gaussian
	Box constraint level	0.001002	0.0011036	0.0013165
	Multiclass method	one-vs-one	one-vs-all	one-vs-all
EDT	Ensemble method	bag	bag	gentleboost
	Maximum number of splits	36	1	50
	Number of learners	11	13	10
	Number of predictors to sample	1	1	1
SDT	Maximum number of splits	1	5	2
	Split criterion	maximum deviance reduction	maximum deviance reduction	maximum deviance reduction

APPENDIX D
ML MODELS' OPTIMIZED HYPERPARAMETERS FOR THE IOWA DATASET

Model	Hyperparameter	70:30	80:20	90:10
NB	Distribution names	Kernel	Kernel	Kernel
	Kernel type	Gaussian	Triangle	Triangle
KNN	Number of neighbours	513	284	323
	Distance metric	Cosine	Hamming	Hamming
	Distance weight	Squared inverse	Equal	Squared inverse
SVM	Kernel function	Gaussian	Gaussian	Linear
	Box constraint level	0.026003	0.0079248	5.5302
	Multiclass method	One-vs-one	One-vs-all	One-vs-one
EDT	Ensemble method	Bag	AdaBoost	logitBoost
	Maximum number of splits	676	1	1
	Number of learners	445	12	12
	Number of predictors to sample	5	1	1
SDT	Maximum number of splits	44	10	10
	Split criterion	maximum deviance reduction	Gini's diversity index	maximum deviance reduction

LIST OF ACRONYMS

Area under the receiver operating characteristics curve	AUC
Bayesian optimization	BO
ensemble decision trees	EDT
Gaussian process	GP
k-nearest neighbors	KNN
Machine learning	ML
Naive Bayes	NB
particle swarm optimization	PSO
Receiver operating characteristic	ROC
single decision trees	SDT
support vector machine	SVM
The multinomial logit	MNL
The out-of-bag	OOB
The US 2017 National Household Travel Survey	NHTS
Work travel mode choice	WTMC

DATA AVAILABILITY
Data used for this study is freely available at https://nhts.ornl.gov/.

REFERENCES
[1] G. Liu, "A behavioral model of work-trip mode choice in Shanghai," *China Econ. Rev.*, vol. 18, no. 4, pp. 456–476, Jan. 2007.
[2] D. Simons, I. De Bourdeaudhuij, P. Clarys, B. de Geus, C. Vandelanotte, J. Van Cauwenberg, and B. Deforche, "Choice of transport mode in emerging adulthood: Differences between secondary school students, studying young adults and working young adults and relations with gender, SES and living environment," *Transp. Res. A, Policy Pract.*, vol. 103, pp. 172–184, Sep. 2017.

[3] D. Badoe, "Modelling work-trip mode choice decisions in two-worker households," *Transp. Planning Technol.*, vol. 25, no. 1, pp. 49–73, Jan. 2002.

[4] N. Day, K. N. Habib, and E. J. Miller, "Analysis of work trip timing and mode choice in the greater Toronto area," *Can. J. Civil Eng.*, vol. 37, no. 5, pp. 695–705, May 2010.

[5] E. Heinen, K. Maat, and B. van Wee, "The effect of work-related factors on the bicycle commute mode choice in The Netherlands," *Transportation*, vol. 40, no. 1, pp. 23–43, Jan. 2013.

[6] M. T. Tran, J. Zhang, M. Chikaraishi, and A. Fujiwara, "A joint analysis of residential location, work location and commuting mode choices in hanoi, Vietnam," *J. Transp. Geography*, vol. 54, pp. 181–193, Jun. 2016.

[7] S. Gandhi and G. Tiwari, "Sociopsychological, instrumental, and sociodemographic determinants of travel mode choice behavior in Delhi, India," *J. Urban Planning Develop.*, vol. 147, no. 3, Sep. 2021, Art. no. 04021028.

[8] Y. Lu, C. G. Prato, N. Sipe, A. Kimpton, and J. Corcoran, "The role of household modality style in first and last mile travel mode choice," *Transp. Res. A, Policy Pract.*, vol. 158, pp. 95–109, Apr. 2022.

[9] C. Ding, X. Cao, B. Yu, and Y. Ju, "Non-linear associations between zonal built environment attributes and transit commuting mode choice accounting for spatial heterogeneity," *Transp. Res. A, Policy Pract.*, vol. 148, pp. 22–35, Jun. 2021.

[10] L. Yang, C. Ding, Y. Ju, and B. Yu, "Driving as a commuting travel mode choice of car owners in urban China: Roles of the built environment," *Cities*, vol. 112, May 2021, Art. no. 103114.

[11] A. Vega and A. Reynolds-Feighan, "Employment sub-centres and travel-to-work mode choice in the Dublin region," *Urban Stud.*, vol. 45, no. 9, pp. 1747–1768, Aug. 2008.

[12] A. Hamre and R. Buehler, "Commuter mode choice and free car parking, public transportation benefits, showers/lockers, and bike parking at work: Evidence from the washington, DC region," *J. Public Transp.*, vol. 17, no. 2, pp. 67–91, Jun. 2014.

[13] Y. Ao, Y. Zhang, Y. Wang, Y. Chen, and L. Yang, "Influences of rural built environment on travel mode choice of rural residents: The case of rural Sichuan," *J. Transp. Geography*, vol. 85, May 2020, Art. no. 102708.

[14] M. Ali, A. R. G. de Azevedo, M. T. Marvila, M. I. Khan, A. M. Memon, F. Masood, N. M. Y. Almahbashi, M. K. Shad, M. A. Khan, R. Fediuk, R. Timokhin, A. Borovkov, and I. U. Haq, "The influence of COVID-19-Induced daily activities on health parameters—A case study in Malaysia," *Sustainability*, vol. 13, no. 13, p. 7465, Jul. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/13/7465

[15] M. Ali, D. B. Dharmowijoyo, I. S. Harahap, A. Puri, and L. E. Tanjung, "Travel behaviour and health: Interaction of activity-travel pattern, travel parameter and physical intensity," *Solid State Technol.*, vol. 63, no. 6, pp. 4026–4039, 2020.

[16] L. Liang, M. Xu, S. Grant-Müller, and L. Mussone, "Household travel mode choice estimation with large-scale data—An empirical analysis based on mobility data in Milan," *Int. J. Sustain. Transp.*, vol. 15, no. 1, pp. 70–85, Jan. 2021.

[17] K. M. Nurul Habib, "Modeling commuting mode choice jointly with work start time and work duration," *Transp. Res. A, Policy Pract.*, vol. 46, no. 1, pp. 33–47, Jan. 2012.

[18] M. Aghaabbasi, Z. A. Shekari, M. Z. Shah, O. Olakunle, D. J. Armaghani, and M. Moeinaddini, "Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques," *Transp. Res. A, Policy Pract.*, vol. 136, pp. 262–281, Jun. 2020.

[19] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Syst. Appl.*, vol. 78, pp. 273–282, Dec. 2017, doi: 10.1016/j.eswa.2017.01.057.

[20] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, P. Zarembka, Ed. New York, NY, USA: Academic, 1973, pp. 105–142.

[21] K. E. Train, *Discrete Choice Methods With Simulation*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[22] M. E. Ben-Akiva, S. R. Lerman, and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA, USA: MIT Press, 1985.

[23] A. Lhéritier, M. Bocamazo, T. Delahaye, and R. Acuna-Agost, "Airline itinerary choice modeling using machine learning," *J. Choice Model.*, vol. 31, pp. 198–209, Jun. 2019.

[24] L. Yang, M. Aghaabbasi, M. Ali, A. Jan, B. Bouallegue, M. F. Javed, and N. M. Salem, "Comparative analysis of the optimized KNN, SVM, and ensemble DT models using Bayesian optimization for predicting pedestrian fatalities: An advance towards realizing the sustainable safety of pedestrians," *Sustainability*, vol. 14, no. 17, p. 10467, Aug. 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/17/10467

[25] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, no. 4. Cham, Switzerland: Springer, 2006.

[26] L. Cheng, X. Chen, J. D. Vos, X. Lai, and F. Witlox, "Applying a random forest method approach to model travel mode choice behavior," *Travel Behav. Soc.*, vol. 14, pp. 1–10, Jan. 2019.

[27] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck, "Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models," *Travel Behav. Soc.*, vol. 20, pp. 22–35, Jul. 2020.

[28] M. Ali, D. B. E. Dharmowijoyo, A. R. G. de Azevedo, R. Fediuk, H. Ahmad, and B. Salah, "Time-use and spatio-temporal variables influence on physical activity intensity, physical and social health of travelers," *Sustainability*, vol. 13, no. 21, p. 12226, Nov. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/21/12226

[29] Y. Chen, M. Aghaabbasi, M. Ali, S. Anciferov, L. Sabitov, S. Chebotarev, K. Nabiullina, E. Sychev, R. Fediuk, and R. Zainol, "Hybrid Bayesian network models to investigate the impact of built environment experience before adulthood on students' tolerable travel time to campus: Towards sustainable commute behavior," *Sustainability*, vol. 14, no. 1, p. 325, Dec. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/14/1/325

[30] P. Salas, R. D. la Fuente, S. Astroza, and J. A. Carrasco, "A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity," *Expert Syst. Appl.*, vol. 193, May 2022, Art. no. 116253.

[31] R. Elshawi, M. Maher, and S. Sakr, "Automated machine learning: State-of-the-art and open challenges," 2019, *arXiv:1906.02287*.

[32] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Cham, Switzerland: Springer, 2013.

[33] T. Ma, M. Aghaabbasi, M. Ali, R. Zainol, A. Jan, A. M. Mohamed, and A. Mohamed, "Nonlinear relationships between vehicle ownership and household travel characteristics and built environment attributes in the U.S. using the XGBT algorithm," *Sustainability*, vol. 14, no. 6, p. 3395, Mar. 2022.

[34] Y. Qian, M. Aghaabbasi, M. Ali, M. Alqurashi, B. Salah, R. Zainol, M. Moeinaddini, and E. E. Hussein, "Classification of imbalanced travel mode choice to work data using adjustable SVM model," *Appl. Sci.*, vol. 11, no. 24, p. 11916, Dec. 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/24/11916

[35] P. Tang, M. Aghaabbasi, M. Ali, A. Jan, A. M. Mohamed, and A. Mohamed, "How sustainable is people's travel to reach public transit stations to go to work? A machine learning approach to reveal complex relationships," *Sustainability*, vol. 14, no. 7, p. 3989, Mar. 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/7/3989

[36] G. I. Diaz, A. Fokoue-Nkoutche, G. Nannicini, and H. Samulowitz, "An effective algorithm for hyperparameter optimization of neural networks," *IBM J. Res. Develop.*, vol. 61, nos. 4–5, pp. 9:1–9:11, Jul. 2017.

[37] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020.

[38] G. Luo, "A review of automatic selection methods for machine learning algorithms and hyper-parameter values," *Netw. Model. Anal. Health Informat. Bioinf.*, vol. 5, no. 1, pp. 1–16, Dec. 2016.

[39] D. Maclaurin, D. Duvenaud, and R. Adams, "Gradient-based hyperparameter optimization through reversible learning," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2113–2122.

[40] N. DeCastro-García, Á. L. M. Castañeda, D. E. García, and M. V. Carriegos, "Effect of the sampling of a dataset in the hyperparameter optimization phase over the efficiency of a machine learning algorithm," *Complexity*, vol. 2019, pp. 1–16, Feb. 2019.

[41] M. Li, W. Du, and F. Nian, "An adaptive particle swarm optimization algorithm based on directed weighted complex network," *Math. Problems Eng.*, vol. 2014, pp. 1–7, Apr. 2014.

[42] W. Xie, W. Nie, P. Saffari, L. F. Robledo, P.-Y. Descote, and W. Jian, "Landslide hazard assessment based on Bayesian optimization-support vector machine in Nanping City, China," *Natural Hazards*, vol. 109, no. 1, pp. 931–948, 2021.

[43] K. Eggensperger, "Towards an empirical foundation for assessing Bayesian optimization of hyperparameters," in *Proc. NIPS Workshop Bayesian Optim. Theory Pract.*, vol. 10, no. 3, pp. 1–5, 2013.

[44] K. Eggensperger, F. Hutter, H. Hoos, and K. Leyton-Brown, "Efficient benchmarking of hyperparameter optimizers via surrogates," in *Proc. AAAI Conf. Artif. Intell.*, vol. 29, no. 1, pp. 1114–1120, 2015.

[45] W. Tao, M. Aghaabbasi, M. Ali, A. H. Almaliki, R. Zainol, A. A. Almaliki, and E. E. Hussein, "An advanced machine learning approach to predicting pedestrian fatality caused by road crashes: A step toward sustainable pedestrian safety," *Sustainability*, vol. 14, no. 4, p. 2436, Feb. 2022.

[46] J. Wang, A. S. Mohammed, E. Macioszek, M. Ali, D. V. Ulrikh, and Q. Fang, "A novel combination of PCA and machine learning techniques to select the most important factors for predicting tunnel construction performance," *Buildings*, vol. 12, no. 7, p. 919, Jun. 2022. [Online]. Available: https://www.mdpi.com/2075-5309/12/7/919

[47] Z. Xu, M. Aghaabbasi, M. Ali, and E. Macioszek, "Targeting sustainable transportation development: The support vector machine and the Bayesian optimization algorithm for classifying household vehicle ownership," *Sustainability*, vol. 14, no. 17, p. 11094, Sep. 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/17/11094

[48] H. Al-Ahmadi, "Development of intercity work mode choice model for Saudi Arabia," *WIT Trans. Built Environ.*, vol. 96, no. 1, pp. 3–21, 2007.

[49] Y. Hatamzadeh, M. Habibian, and A. Khodaii, "Walking mode choice across genders for purposes of work and shopping: A case study of an Iranian city," *Int. J. Sustain. Transp.*, vol. 14, no. 5, pp. 389–402, May 2020.

[50] E. Heinen and W. Bohte, "Multimodal commuting to work by public transport and bicycle: Attitudes toward mode choice," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2468, no. 1, pp. 111–122, Jan. 2014.

[51] S. Indriany, A. Sjafruddin, A. Kusumawati, and W. Weningtyas, "Mode choice model for working trip under risk and uncertainty," *AIP Conf. Proc.*, vol. 1977, no. 1, 2018, Art. no. 020041.

[52] M. Irfan, A. N. Khurshid, M. B. Khurshid, Y. Ali, and A. Khattak, "Policy implications of work-trip mode choice using econometric modeling," *J. Transp. Eng., A, Syst.*, vol. 144, no. 8, Aug. 2018, Art. no. 04018035.

[53] P. Kunhikrishnan and K. K. Srinivasan, "Choice set variability and contextual heterogeneity in work trip mode choice in Chennai city," *Transp. Lett.*, vol. 11, no. 4, pp. 174–189, Jul. 2019.

[54] V. Vapnik, *The Nature of Support Vector Machine*. Berlin, Germany: Springer, 1999.

[55] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[56] D. J. Hand and K. Yu, "Idiot's Bayes? Not so stupid after all?" *Int. Stat. Rev.*, vol. 69, no. 3, pp. 385–398, Dec. 2001.

[57] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[58] C. Xie, J. Lu, and E. Parkany, "Work travel mode choice modeling with data mining: Decision trees and neural networks," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1854, no. 1, pp. 50–61, Jan. 2003.

[59] M. T. Kashifi, A. Jamal, M. S. Kashefi, M. Almoshaogeh, and S. M. Rahman, "Predicting the travel mode choice with interpretable machine learning techniques: A comparative study," *Travel Behav. Soc.*, vol. 29, pp. 279–296, Oct. 2022.

[60] S. F. Franco, "Downtown parking supply, work-trip mode choice and urban spatial structure," *Transp. Res. B, Methodol.*, vol. 101, pp. 107–122, Jul. 2017.

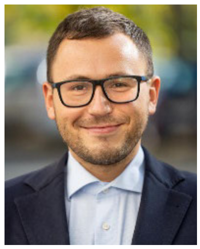
[61] Y. Lu and K. Kawamura, "Data-mining approach to work trip mode choice analysis in Chicago, Illinois, area," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2156, no. 1, pp. 73–80, Jan. 2010.

[62] Z. Patterson, G. Ewing, and M. Haider, "Gender-based analysis of work trip mode choice of commuters in suburban montreal, canada, with stated preference data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1924, no. 1, pp. 85–93, Jan. 2005.

[63] A. Vega and A. Reynolds-Feighan, "A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns," *Transp. Res. A, Policy Pract.*, vol. 43, no. 4, pp. 401–419, May 2009.

MAHDI AGHAABBASI is currently a Researcher with the Transportation Institute, Chulalongkorn University. His research interests include transportation planning, the application of machine learning for solving transportation issues, assessing travel behavior changes, and revealing complex relationships between the built environment and travel behaviors.

MUJAHID ALI is currently pursuing the Ph.D. degree with the Department of Transport Systems, Traffic Engineering, and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology. His research interests include transportation engineering, travel behavior, and health.



MICHAŁ JASIŃSKI (Member, IEEE) received the Ph.D. and D.Sc. degrees in electrical engineering from the Wrocław University of Science and Technology, in 2019 and 2022, respectively. Since 2018, he has been with the Electrical Engineering Faculty, Wrocław University of Technology, where he is currently an Assistant Professor. He is the author and coauthor of over 100 scientific publications. His research interests include using big data in energy systems and multidisciplinary applications of machine learning. He is also the Guest Editor of Special Issues in Energies, Electronics, Sustainability and *Frontiers in Energy Research*.

ZBIGNIEW LEONOWICZ (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Wrocław University of Science and Technology, in 1997 and 2001, respectively, and the Habilitation degree from the Bialystok University of Technology, in 2012. Since 1997, he has been with the Electrical Engineering Faculty, Wrocław University of Technology. He also received the two titles of a Full Professor from the President of Poland and the President of the Czech Republic, in 2019. Since 2019, he has been a Professor with the Department of Electrical Engineering. He is currently the Head of the Chair of Electrical Engineering Fundamentals with the Department of Electrical Engineering.

TOMÁŠ NOVÁK currently works with the Department of Electrical Power Engineering, Faculty of Electrical Engineering and Computer Science, VSB—Technical University of Ostrava, as an Associate Professor. He is also a lecturer, a researcher, and a supervisor to Ph.D. students for the problems of lighting technology. His current research interests include public lighting, light pollution, interior light controlling, and smart city technologies. He has experience with technical designing and measurement of lighting systems. He is the Chair of Czech Lighting Society, a member of Czech National Committee of the CIE and a member of CIE TC 4-58.