Article

Understanding Active Transportation to School Behavior in Socioeconomically Disadvantaged Communities: A Machine Learning and SHAP Analysis Approach

Bita Etaati [1], Arash Jahangiri [2,*], Gabriela Fernandez [3], Ming-Hsiang Tsou [3] and Sahar Ghanipoor Machiani [2]

[1] Big Data Analytics (BDA) Program, San Diego State University, San Diego, CA 92182, USA; bita.et@gmail.com
[2] Department of Civil, Construction, and Environmental Engineering, San Diego State University, San Diego, CA 92182, USA; sghanipoor@sdsu.edu
[3] Department of Geography, San Diego State University, San Diego, CA 92182, USA; gfernandez2@sdsu.edu (G.F.); mtsou@sdsu.edu (M.-H.T.)
* Correspondence: ajahangiri@sdsu.edu

Abstract: Active Transportation to School (ATS) offers numerous health benefits and is considered an affordable option, especially in disadvantaged neighborhoods. The US Centers for Disease Control and Prevention (CDC) advises 60 min of daily physical exercise for children aged 6 to 17, making ATS a compelling approach to promote a healthier lifestyle among students. Initiated in 2005 by the US Department of Transportation (DOT), the Safe Routes to School (SRTS) program aims to foster safe and regular walking and biking to school for students. This paper examines students' travel behavior using SRTS survey data and assesses the program's effectiveness in promoting ATS in Chula Vista, California. Employing machine learning algorithms (random forest, logistic regression, and support vector machines) to predict students' likelihood to walk to school, it utilizes SHAP (SHapley Additive exPlanations) to pinpoint significant variables influencing ATS across all models. SHAP underscores critical factors affecting transportation choices to school, highlighting the importance of home-to-school distance, with shorter distances positively impacting active transportation. However, only half of students within schools' walking distance opted to walk to school, underscoring the necessity of addressing parental safety concerns, including factors such as crime rates and traffic speed along the route.

Keywords: Safe Routes to School; logistic regression; support vector machines; random forest; SHapley Additive exPlanations

1. Introduction

Engaging in physical activity during adolescence offers a range of advantages for physical, social, and psychological well-being [1–3]. Therefore, the Centers for Disease Control and Prevention (CDC) advise that children and teenagers should participate in a minimum of 60 min of daily physical activity [4]. However, studies indicate that only about one-third of children actually meet this recommendation. This concerning lack of sufficient physical activity among youth has become a prominent concern in the public health community [5].

Consistent exercise is crucial for the physical and emotional health of children, particularly for those battling obesity. Children and adolescents who participate in physical activity are less likely to become overweight and obese during their youth and adolescence, and they are less likely to become obese as adults [6,7]. One potential factor that may contribute to increased physical activity and reduced childhood obesity is active commuting to school [8–11]. Surprisingly, when comparing data from 1969 to 2009, there has been a significant decline in the number of students who walk or cycle to school. In less than 40 years, the average of students walking or cycling to school has dropped from 42% to

13%. Statistics show that in 1969, 87% of those who lived within one mile of school would walk or bicycle to school, and this number decreased to almost half (47%) in 2009 [12].

This paper centers on Chula Vista, San Diego County's second-largest city, which houses a significant number of school-aged children. With approximately 270,000 residents, a quarter of whom are children under the age of eighteen, the city has around fifty-two elementary schools. Alarmingly, 38% of school-aged children in Chula Vista were reported as overweight/obese by 2010 [13]. Highlighting the city's challenges, the Social Vulnerability Index (SVI) from the Centers for Disease Control and Prevention (CDC) and the Agency for Toxic Substances and Disease Registry (ATSDR) ranks Chula Vista among the cities with the highest social vulnerability in San Diego County [14]. This vulnerability is gauged through various metrics such as poverty and vehicle access, among others.

In an effort to promote Active Transportation to School (ATS), the Safe Routes to School (SRTS) program was introduced. SRTS is a federally funded program which encourages students to walk and cycle to school through 6 Es: engagement, equity, engineering, encouragement, education, and evaluation. This program focuses on both infrastructure improvements and educational programs and hopes to foster a culture of active transportation among students. Since 2005, more than 14,000 schools all across the US have participated in this program [15]. The Chula Vista Elementary School District (CVESD) has been an active SRTS participant since 2007. Numerous activities and projects were initiated by SRTS in CVESD, including pedestrian safety educational programs and bicycle rodeos.

Surveys have proven to be a valuable and cost-effective method for collecting information from a large population [16]. Consequently, the National Center for Safe Routes to School program has provided standardized surveys, including the Parent survey, which seeks information about students' modes of transport, factors influencing parental decisions regarding their child's commute, safety conditions along routes to school, and other relevant background information. These surveys can play a crucial role in identifying barriers to active transportation and measuring changes in parental attitudes as a result of local SRTS programs. To evaluate the effectiveness of SRTS programs on active transportation trends among CVESD students, responses to the Parent survey were collected before, during, and after the implementation of SRTS activities.

This research aims to explore the factors that hinder students from engaging in Active Transportation to School, with a particular focus on understanding the perspectives of parents, the role of schools in promoting active transportation, and the influence of students' home distance from school. By thoroughly investigating these barriers, we aspire to uncover valuable insights that can pave the way for effective interventions and initiatives to encourage active transportation among students.

Furthermore, this study seeks to assess the impact of the SRTS activities on the promotion of active transportation. We rigorously analyzed whether the implementation of SRTS initiatives leads to a significant increase in the percentage of students utilizing active modes of transportation, such as walking or cycling, to commute to school. By examining data from various stages of the program, including pre-implementation, mid-implementation, and post-implementation, this study provides a comprehensive understanding of the program's effectiveness over time.

Following an extensive literature review on the subject matter, the authors of this study utilized multiple machine learning algorithms. Leveraging data collected by the National Center for Safe Routes to School, we employed advanced statistical techniques and predictive modeling to pinpoint key factors that influence students' transportation choices.

The findings of this research have the potential to generate promising outcomes and inform evidence-based strategies to overcome barriers to active transportation among students. By thoroughly analyzing the Parent survey, we can develop targeted interventions that promote and sustain active transportation behaviors. Ultimately, this study contributes to the development of practical and impactful approaches that will empower more students to embrace active transportation and lead healthier, more active lives.

2. Previous Work

Many of the previous studies on the Safe Routes to School primarily focus on changes in travel behavior associated with SRTS program interventions. These studies have predominantly highlighted the impact of infrastructure improvements on ATS [17–19]. While there may be a positive correlation between traffic improvements and an increase in ATS rates, it should be noted that infrastructure enhancements alone might not be sufficient to promote cycling and walking in schools with low levels of activity [20]. In fact, previous research suggests that non-infrastructure measures, such as educational programs, can effectively encourage walking and biking to school [21–23].

In addition to infrastructural and non-infrastructural factors, there are several elements that influence the likelihood of engaging in active modes of transportation when traveling to and from school. Previous research has emphasized the significance of home-to-school distance in determining the rate of active transportation among children and adolescents [24,25]. Notably, a greater home-to-school distance between the residential area and the school has been identified as a robust predictor of active transportation behavior in children and adolescents [26–28]. Increased distances between home and school often result in children traveling on arterial roads and having to cross them, which poses significant traffic-related safety challenges [29]. Recent studies have demonstrated that parents' attitudes toward Active Transportation to School have become more negative as the distance between their homes and schools has increased [30].

In addition to the home-to-school distance, mediating factors can also influence students' decision to use ATS [31]. These mediating factors encompass the direct influence of the social and natural environment, such as crime rates and traffic collision rates, as well as the perceptual interpretations of parents and children regarding their surroundings, including perceived risks associated with crime or traffic. In some locales, regional policies and land development also significantly shape these active transportation choices [32]. A thorough examination of over 60 papers on active travel has revealed the impact of mediating factors such as age and education levels in travel preferences. Some of the previous studies suggest that active travel decreases with age, with men and those with higher education preferring biking, while women tend to choose walking. Moreover, these studies consistently note lower bike usage among minorities and those with lower incomes [33].

Previous research by Davison et al. suggests that parental perceptions of the environment have a stronger impact on transportation patterns to school than built environment factors alone [11]. Parents' concerns about traffic and safety related to ATS are significant factors contributing to the increasing number of parents choosing to drive their children to and from school [8,34]. Previous studies indicate that parental barriers to engaging in active commuting are shaped by various factors, including the age, gender, mode of transportation employed by their children [35], and parents' level of education [36–38]. Parents' concerns primarily arise from the absence of sidewalks, the presence of heavy traffic, high-speed roads, risky pedestrian crossings, and personal safety issues such as crime, collectively shaping their perception of unsafe routes to school [39–42]. The concerns of parents regarding traffic, crime, and personal safety can significantly influence the choice of transportation mode to school for adolescents, consequently shaping their perceptions of the route and their behavior regarding Active Transportation to School [30,43,44]. Specifically, personal safety concerns, such as local crime rates and the presence of strangers in the neighborhood, also discourage children from engaging in ATS and opting to walk or cycle to school [45,46]. Moreover, parental concerns about the safety of their children while walking and cycling to school are further compounded by the lack of adult supervision [47,48].

Barriers to active transportation can have a significant impact on participation rates. Studies have shown that children who do not face reported barriers are more likely to walk or bike to school compared to their peers who encounter one or more barriers [49]. Therefore, it is crucial to address traffic and personal safety concerns when developing models for safe walking and cycling routes to schools [50]. By considering these findings, efforts can be made to create frameworks that address these barriers, ultimately promoting

and encouraging active transportation among children and adolescents. By addressing these barriers, we can further enhance our understanding of the factors influencing active transportation to school and develop evidence-based strategies that promote safe and sustainable travel options for children and adolescents. Such efforts can contribute to creating healthier, more livable communities where Active Transportation to School becomes the norm rather than the exception.

In the landscape of school travel mode choice modeling, logistic regression (LR) has remained a time-tested and trusted tool. Numerous studies have explored its capabilities; for instance, a study spanning from 2007 to 2018 in Arizona employed LR to identify students' chosen modes of transport to school [51]. Another notable investigation utilized it to evaluate the challenges faced by student commuters, considering diverse demographic and institutional factors [52]. Such varied applications underscore LR's essential role in educational inquiries. Moving into more recent analytical advancements, the random forest (RF) technique is gaining ground in transportation studies. Previous work [53–55] serves as proof of RF's capability to navigate and interpret complex traveler behaviors. Further solidifying its value, one previous study's [56] exploration using RF on Nanjing's travel diary data emphasized the method's strength in both enhancing prediction accuracy and analyzing travel determinants. Lastly, support vector machines (SVM) have also demonstrated significant efficacy in this field. Dave et al.'s study on the transportation preferences of Vadodara's schoolchildren and Assi, Khaled J., et al.'s innovative blend of SVM with clustering techniques to forecast student travel mode decisions highlight SVM' pivotal importance [57,58]. Together, LR, RF, and SVM offer a comprehensive analytical framework, each contributing uniquely to our understanding and prediction of school travel mode choices.

Given the SRTS initiatives in Chula Vista, a prior study examined students' transportation habits and underscored the importance of factors such as school proximity, crime concerns, and school encouragement in shaping student decisions [59]. Building on this research, the current study aims to extend the existing literature by conducting an analysis of survey data gathered from parents at distinct stages—pre-implementation, mid-implementation, and post-implementation—of these active transportation initiatives. With this approach, we seek to provide a deeper understanding of evolving parental concerns and perceptions as these interventions unfold over time.

3. Methodology

3.1. Data Collection

The surveys created by the National Center for Safe Routes to School were studied to provide a better insight into the school travel environment. The Parent survey proposed by this center gathers information from parents/guardians on their children's travel behavior, including their usual transportation mode and how far they live from their school, and the concerns parents may have for active transportation. This survey mainly focuses on the issues that may affect parents' willingness or permission for their children to walk/bicycle to school. The Parent survey also includes children's background information (age, gender), and if parents are concerned about twelve potential issues mentioned in the survey. This can be used as a powerful tool to investigate the underlying reasons why students are (or are not) considering active school travel, and if these issues arise from safety concerns or parents' perceptions. The Parent dataset used in this research was gathered and consolidated for all SRTS participant schools in CVESD, and included 5764 surveys collected from 19 schools between 2009 and 2011, for students aged from Pre-K to sixth grade. Table 1 presents a detailed overview of the features derived from the Parent Survey, along with their respective explanations.

Table 1. Features Derived from SRTS Parent Survey.

Feature	Explanation
Transportation Mode	**Response Variable**. Indicates the transportation mode of children when they leave school (Transportation Mode = 0 (do not walk to school) or Transportation Mode = 1 (walk to school)).
Period	The period of the Safe Routes to School (SRTS) program: before, mid, or after implementing the SRTS program at the school.
Child_Gender	The gender of the child.
Distance_from_School	The distance of the child's home from the school (less than 1/4 of a mile, 1/4 of a mile up to 1/2 a mile, 1/2 a mile up to 1 mile, 1 mile up to 2 miles, more than 2 mile).
child_asked_permission	Indicates whether the child asked for permission to walk/cycle to school (yes/no).
grade_allowed	The grade in which the child is allowed to use ATS by their parents.
issue_distance	Concerns about the ATS distance affecting parents' decision on their child's school commute (yes/no).
issue_convienience	Concerns about the convenience of ATS affecting parents' decision on their child's school commute (yes/no).
issue_time	Concerns about the time required for ATS affecting parents' decision on their child's school commute (yes/no).
issue_after_school_program	Concerns about the after-school programs affecting parents' decision on their child's school commute (yes/no).
issue_speed	Concerns about the speed limit affecting parents' decision on their child's school commute (yes/no).
issue_traffic	Concerns about the traffic affecting parents' decision on their child's school commute (yes/no).
issue_walk_with_adults	Concerns about children walking without adult supervision affecting parents' decision on their child's school commute (yes/no).
issue_side1	Concerns about safety of sidewalks affecting parents' decision on their child's school commute (yes/no).
issue_intersection	Concerns about safety of intersections affecting parents' decision on their child's school commute (yes/no).
issue_Crossing_Guards	Concerns about safety of crossing guards affecting parents' decision on their child's school commute (yes/no).
issue_crime	Concerns about crime rates affecting parents' decision on their child's school commute (yes/no).
issue_weather	Concerns about the weather affecting parents' decision on their child's school commute (yes/no).
school_encouragement	Indicates whether the school encourages walking or biking to school (encourages, neither, discourages).
child_having_fun	Represents whether the child finds walking or biking to school fun (fun, neutral, boring).
healthy	Indicates whether parents believe ATS is healthy or safe for the child (health, neutral, unhealthy).
Parent_education	Education level of the parent(s) (elementary, some high school, high school graduate, some college, college graduate).

According to the American Academy of Pediatrics (AAP) [60] and previous studies [61,62], it is generally recommended that children wait until they are in fifth grade, around the age of 10, before walking to school without adult supervision. Hence, this study only focused on the data from 5th to 12th grade students. After removing the irrelevant student age groups (Pre-K to fourth grade) and some missing values, the data were reduced to 1387 observations.

3.2. Statistical Analysis

The primary objective of this research was to investigate the factors influencing the choice of transportation mode for students traveling to school, focusing on active transporta-

tion. Due to the limited number of observations for bicycling, the analysis was narrowed down to solely examine walking. However, it should be noted that the same analytical framework can be readily applied to bicycling data if a sufficient volume of observations becomes available.

We extracted independent variables from the Parent survey and used them alongside transportation mode as the dependent variable in our analysis. Transportation mode indicated in the Parent survey was transformed into a binary classification format, indicating whether the student walks as their usual commuting mode (Transportation Mode = 1) or if they use any other transportation mode (Transportation Mode = 0).

The independent variables used in the analysis encompassed factors such as students' gender, distance between home and school, the role of the school in promoting active transportation, as well as parents' education level, perceptions, and concerns regarding active transportation.

To identify the key factors influencing the decision to walk to school, we employed multiple machine learning algorithms, namely logistic regression (LR), random forest (RF), and support vector machines (SVM). Given the potential issue of multicollinearity arising from highly correlated independent variables [63,64], we performed a preliminary analysis to identify and remove such variables. Additionally, to mitigate overfitting concerns and obtain more accurate estimates, we applied 5-fold cross-validation to each model [65,66]. Finally, SHapley Additive exPlanations values (SHAP) were used as a means of feature selection and compared across the models to identify the most significant factors influencing the prevalence of walking to school.

It is worth noting that the Parent survey provided options for two transportation modes: morning (going to school) and evening (coming back home). One of the objectives of this research was to compare these responses and determine if significant differences existed in the patterns of walking to and from school. Accordingly, two logistic regression models were constructed for this comparison. The results did not show significant variations, with a slightly higher prevalence of students walking back home in the evening (27%) compared to walking to school in the morning (23%). Consequently, this paper only used the morning transportation mode as the response variable in all three models.

3.3. Model Selection

3.3.1. Logistic Regression

Regression methods have emerged as a fundamental component of data analysis in exploring the connection between a dependent variable and one or multiple independent variables. Logistic regression, a statistical model commonly employed in traffic safety studies [67–71], has been utilized for investigating the association between a binary response variable and independent variables.

Logistic regression is a powerful statistical tool optimized for predicting binary outcomes based on one or more explanatory variables. Unlike linear regression, which predicts continuous values, logistic regression focuses on estimating the probability that a given observation falls into a specific category. Central to this method is the logistic function, which constrains predicted probabilities to lie between 0 and 1 [72].

Mathematically, for predictor variables denoted as $X_1, X_2, \ldots X_n$, the probability of the desired outcome is modeled as:

$$P(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n}}$$

In this formulation, the term $P(X)$ indicates the likelihood of the event in question. The coefficients $(\beta_0, \beta_1, \ldots \beta_n)$ represent the influence of each predictor variable on the log odds of the outcome. Specifically, a coefficient reveals how the log odds of the outcome change with a one-unit increase in its associated predictor, while keeping all other predictors constant. The values of these coefficients are determined using the Maximum Likelihood

Estimation (MLE) method. This method aims to find the coefficient values that are most likely to produce the observed data, given the model's structure.

3.3.2. Random Forest

Random forest (RF) is an ensemble model that effectively leverages decision trees to handle complex, nonlinear relationships and high-dimensional variables, while exhibiting robustness against outliers and noise [73]. The essence of RF lies in its bootstrapping technique and aggregation method. For each tree, a subset of data is sampled with replacement (bootstrap sample), and a subset of features is chosen randomly to split the nodes. This diversification ensures that individual trees capture different patterns in the data. The final prediction, for classification, is based on a majority vote, and for regression, it is the average of the predictions from all trees [72]. Given its capabilities, RF has garnered considerable recognition and has been widely applied in various transportation research contexts [74–78].

For a classification problem, given a new observation *X*, the *RF* prediction is:

$$RF(X) = mode\{T_1(X), T_2(X), ..., T_n(X)\}$$

where $T_i(X)$ is the prediction of the *ith* tree. For regression, it is the average:

$$RF(X) = \frac{1}{n}\sum_{i=1}^{n} T_i(X)$$

The efficacy of random forest (*RF*) models is substantially impacted by the configuration of hyperparameters. To maximize the performance of *RF*, it is imperative to identify the most suitable parameter values through careful optimization. By emphasizing the objective of minimizing the out-of-bag (OOB) error and identifying the optimal number of trees for the *RF* model, the development of the optimal random forest model was accomplished following a 5-fold cross-validation procedure.

3.3.3. Support Vector Machines

Support vector machines (SVM) are widely recognized as one of the highly effective algorithms for classification and regression problems. Due to their extensive application and reliable performance across various scientific domains, SVM have been a focal point in transportation research in recent years [79–81]. One prominent application of this model is for binary classification tasks, aiming to identify the optimal hyperplane that effectively partitions the data into two distinct classes [82,83].

SVM aim to find a hyperplane defined by *w* (weight vector) and *b* (bias) that maximizes the margin between two classes. This margin represents the distance between the hyperplane and the nearest data points, or "support vectors", from both classes. Given labeled data (x_i, y_i) where $y_i \in \{-1, 1\}$, the decision function is defined as:

$$f(X) = w^T x + b$$

The primary goal is to optimize:

$$min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i} \xi_i$$

Here, the objective is to strike a balance. The term $\frac{1}{2}||w||^2$ seeks the hyperplane with the largest possible margin, while $C\sum_{i} \xi_i$ allows for some flexibility, permitting certain points to be on the "wrong side" of the hyperplane for the sake of better overall fit. The parameter "*C*" determines this balance: higher values stress the importance of each data point being correctly classified, even if it means a smaller margin.

The SVM model was developed using scikit-learn. In addition, the effectiveness of SVM can be substantially enhanced by employing parameter optimization techniques [84,85]. In this case, the grid search technique is employed to determine the best values for the regularization parameter "C" and the radial basis kernel parameter "Gamma". The grid search is performed within the context of a 5-fold cross-validation, ensuring reliable evaluation of the model's performance.

3.4. Model Evaluation Metric

Accuracy and error rate are commonly used metrics for evaluating classification models in both binary and multi-class problems due to their simplicity, applicability to different scenarios, and ease of interpretation. However, these metrics have limitations in terms of producing less distinct and discriminative values and showing bias towards majority class instances [86,87]. To overcome these limitations, it is necessary to consider other evaluation metrics.

One widely used evaluation metric is the confusion matrix, which provides a comprehensive assessment of a classifier's performance by capturing true positive, true negative, false positive, and false negative outcomes [88]. The AUC (Area Under the ROC Curve) is another model evaluation metric for assessing classification model performance [89,90]. A value of 1 indicates an ideal model that can accurately separate positive and negative classes, while 0.5 indicates that it performs no better than random.

In this study, the evaluation of the proposed models included the assessment of AUC and confusion matrix, complementing the analysis based on accuracy and error rate. These diverse metrics provide a more comprehensive understanding of the model's performance and contribute to a thorough evaluation process.

3.5. SHAP Values

SHapley Additive exPlanations, often referred to as SHAP, is a method used in machine learning to determine the impact of individual features on a model's predictions [91,92]. By evaluating the contribution of each feature in the dataset to the overall output and considering all possible feature combinations, SHAP values provide insights into how different features affect the model's predictions. Mathematically, the SHAP value of a feature *j* for a specific instance *x* can be expressed as:

$$\phi_j\left(x\right) = \sum\nolimits_{s \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f_x\left(S \cup \{j\}\right) - f_x\left(S\right) \right]$$

where *N* is the set of all features, *S* is a subset of *N* excluding feature *j*, and $f_x(S)$ represents the model's prediction when only features in subset *S* are considered [93]. This formula quantifies the marginal contribution of feature *j* to the model's prediction, averaged over all possible subsets of features, thereby providing an assessment of its pertinence relative to a baseline that takes into account all feature interactions.

This approach aids in understanding the contribution of each feature to the prediction result and can be used as a feature selection mechanism [93–95]. In this study, SHAP values were utilized to determine and compare the most significant factors in ATS for each model. By comparing the SHAP values of each feature in the models, we identified the features that have the most substantial impact on the prediction outcomes.

In binary classification, SHAP values generally range between −1 and +1. These values represent a feature's influence on the model's output compared to the baseline prediction, often derived from the mean prediction of the dataset. A SHAP value in the positive domain suggests that a particular feature drives the model's prediction toward the positive class, while a value in the negative domain indicates a push toward the negative class. The absolute magnitude of the SHAP value, irrespective of its sign, indicates the intensity of the feature's effect on the model's predictive outcome.

4. Results

As discussed earlier in the methodology section, the proposed machine learning models, namely logistic regression (LR), random forest (RF), and support vector machines (SVM), were employed to investigate the relationship between students' transportation mode and various demographic and perceptual features extracted from the Parent survey. The objective was to predict the transportation mode using binary classification, with "walk" assigned a value of 1 and other transportation modes assigned a value of 0.

To enhance the performance of the models, hyperparameter tuning was conducted using Scikit Learn's GridSearchCV method for LR and SVM, and RandomizedSearchCV for RF. GridSearchCV systematically explored a predefined grid of hyperparameter values, evaluating the effectiveness of the models through cross-validation. The optimal hyperparameter configuration was determined based on the highest performance score. RandomizedSearchCV, on the other hand, randomly sampled a subset of hyperparameter combinations from a specified distribution, reducing computational costs while still evaluating performance through cross-validation.

We opted for GridSearchCV for LR and SVM due to their relatively constrained hyperparameter tuning needs: LR is predominantly concerned with regularization aspects, while SVM concentrate on kernel choices. In contrast, RF presents a richer hyperparameter spectrum, encompassing decisions such as optimal tree count (Figure 1), tree depth, and the necessary sample count for splitting internal nodes. Given this complexity, we opted for RandomizedSearchCV, which provides detailed tuning without the extensive demands of a complete grid search.
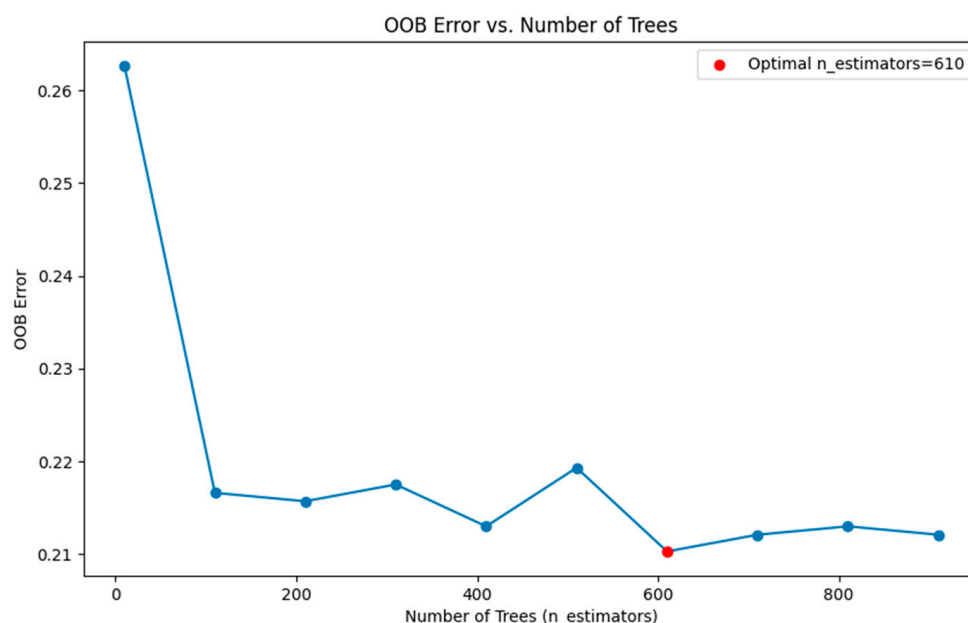

Figure 1. Optimal Number of Trees in RF Model Based on OOB Error.

Employing a 5-fold cross-validation technique for each model, both algorithms facilitated the selection of the most favorable hyperparameter configuration (Tables 2–4), thereby improving the overall performance of the model.

Table 2. Hyperparameters for LR Model.

Hyperparameter	Value	Explanation
'c'	0.1	Regularization Strength (C): A smaller C value (e.g., 0.1) signifies stronger regularization, promoting a preference for simpler model parameter values. This helps the model generalize effectively to unseen data and mitigates the risk of overfitting.
'class_weight'	None	Determines the balance between classes. Determines how classes are weighted in the model. When set to 'None', it indicates equal weighting, signifying that all classes are treated equally in the model.
'penalty'	L2 (Ridge)	Regularization Type: Ridge regularization introduces a penalty term based on the square of the model's coefficient values into the loss function. This regularization method encourages the model to generalize better by constraining the magnitudes of its parameters.
'solver'	Liblinear	Optimization Algorithm used to find the best model coefficients. The 'Liblinear' optimization algorithm is well-suited for small- to medium-sized datasets and pairs effectively with 'L2' regularization.

Table 3. Hyperparameters for RF Model.

Hyperparameter	Value	Explanation
'bootstrap'	True	Controls the use of random sampling with replacement when constructing each tree in the forest. 'True' enables bootstrapping, adding diversity to the ensemble.
'criterion'	Gini	Specifies the rule for splitting tree nodes. Criterion 'Gini' measures data impurity in nodes for better classification by minimizing impurity during splits.
'max_depth'	15	Defines the maximum depth of each tree in the forest. A depth of 15 means trees are limited to 15 levels.
'max_features'	Log2	This hyperparameter controls the maximum number of features considered when making each split in a decision tree. 'Log2' considers a logarithmic number of features per split.
'n_estimators'	610	Specifies the number of trees in the random forest. The value of 610 was chosen based on an analysis of the Out-of-Bag (OOB) error (Figure 1), a metric used to assess model performance. The selection aims to strike a balance between model complexity and generalization.

Table 4. Hyperparameters for SVM Model.

Hyperparameter	Value	Explanation
'c'	10	Regularization Strength (C): A higher C value (e.g., 10) signifies weaker regularization, allowing the SVM to fit the training data more closely. It might increase the risk of overfitting but can capture complex patterns in the data.
'degree'	2	Degree of Polynomial Kernel: The 'degree' hyperparameter sets the degree of the polynomial kernel function. In this case, '2' represents a quadratic kernel, which can capture non-linear relationships in the data.
'gamma'	0.01	Kernel Coefficient (Gamma): 'Gamma' controls the shape of the decision boundary. A lower value (e.g., 0.01) makes the boundary more spread out, potentially leading to smoother decision boundaries.

To evaluate the models' performance and select the best-performing model, several evaluation methods were considered. In addition to the accuracy metric (Table 5), the Performance Metrics (Tables 6–8) and AUC-ROC curve (Figure 2) were analyzed for each model.

Table 5. Model Accuracy.

Model	Accuracy
LR	77%
RF	**80%**
SVM	**80%**

Table 6. Performance Metrics for LR Model.

LR	Precision	Recall	F1-Score
Walk = 0	0.82	84%	0.83
Walk = 1	0.67	63%	0.65

Table 7. Performance Metrics for RF Model.

RF	Precision	Recall	F1-Score
Walk = 0	0.84	85%	0.84
Walk = 1	0.7	67%	0.68

Table 8. Performance Metrics for SVM Model.

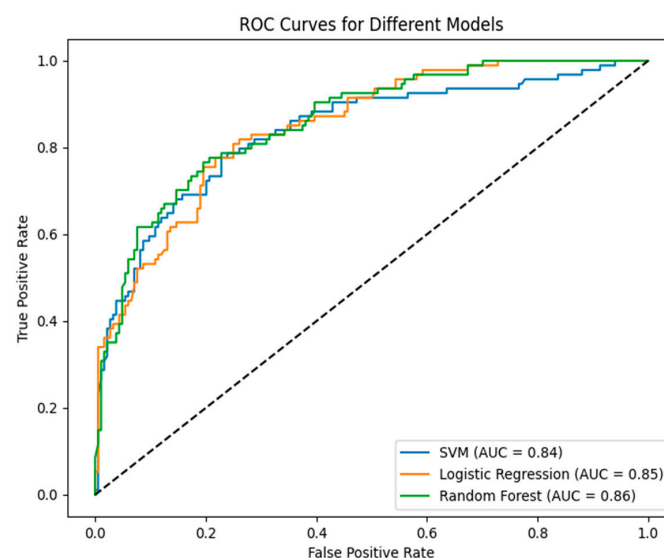| SVM	Precision	Recall	F1-Score
Walk = 0	0.84	**86%**	0.85
Walk = 1	0.71	**67%**	0.69

Figure 2. AUC–ROC curve for SVM, LR, and RF.

The comparative analysis of evaluation metric results for the three models reveals a close performance proximity in terms of essential metrics, including AUC, precision, recall, and F1-score. This observation suggests that all three models demonstrate proficient predictive capabilities for the binary classification task of discerning whether a child will walk home from school or not. Notably, SVM and RF exhibit marginally higher accuracy than the LR model. Hence, it can be inferred that these two models, SVM and RF, outperform the LR model in this context.

To gain deeper insights into each models' performance and understand the impact of individual features on ATS, permutation importance (SHAP values) was employed (Figures 3–5). By measuring the decrease in model performance when randomly shuffling the values of a particular feature while keeping others unchanged, permutation importance allowed us to identify the features that significantly influenced the model's predictions [96].
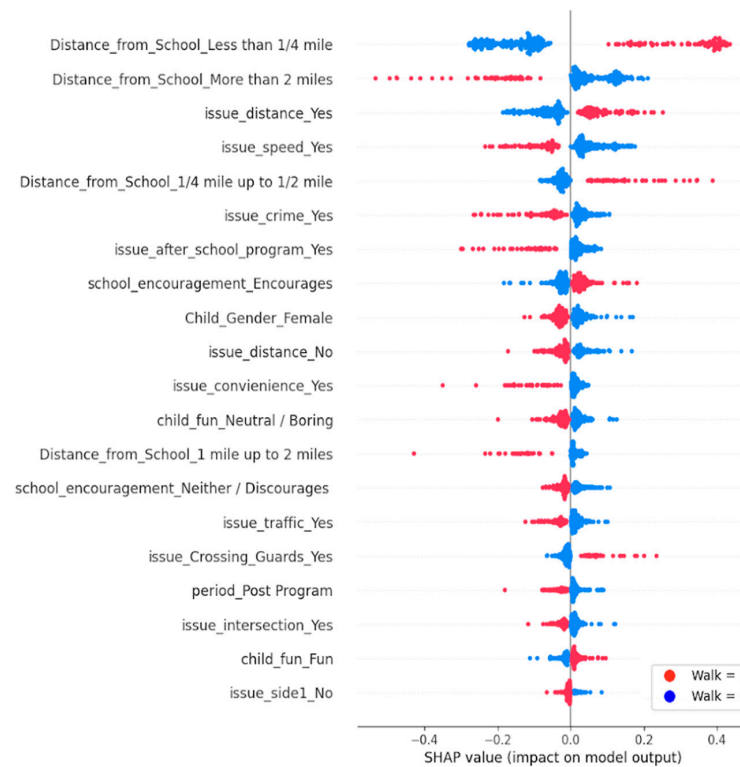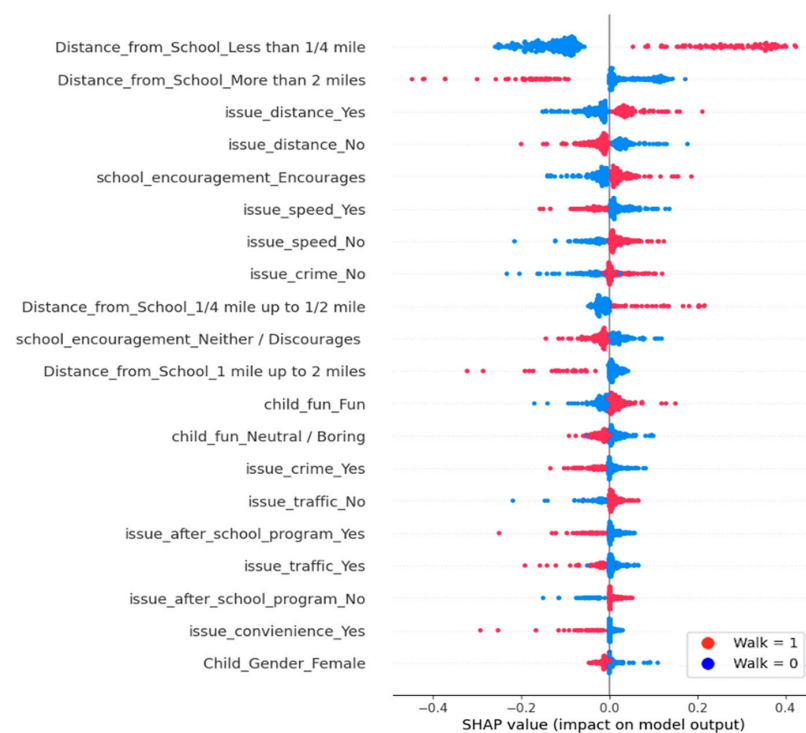

Figure 3. SHAP plot for LR.

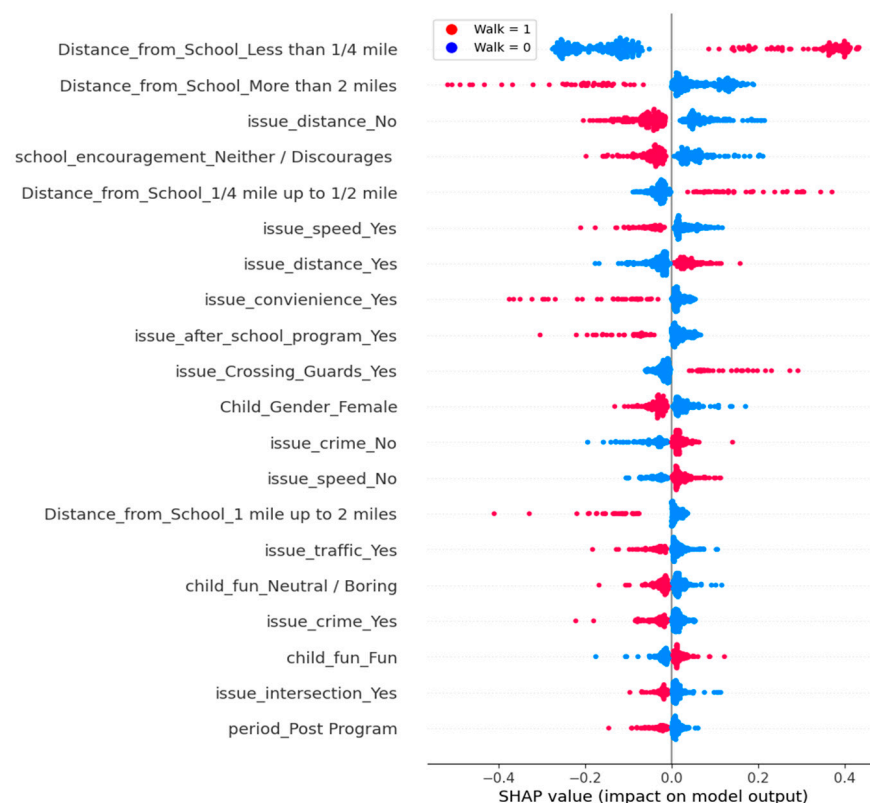**Figure 4.** SHAP plot for RF.

Figure 5. SHAP plot for SVM.

5. Discussion

Utilizing data from the Safe Routes to School program, this study aims to uncover barriers to active transportation among students in the Chula Vista Elementary School District. The comprehensive analysis of SHAP values for the support vector machines (SVM), random forest (RF), and linear regression (LR) models has provided valuable insights into the factors influencing transportation choices to school. Our study findings align with previous research that has primarily focused on changes in travel behavior associated with SRTS program interventions. The analysis through SHAP values consistently demonstrated similar performances and feature ranking across the three machine learning models.

Aligning with previous studies, our study highlights the crucial role of home-to-school distance in shaping students' transportation choices. Children and adolescents are more likely to engage in active transportation when the distance between their residential area and school is less than 1/4 of a mile or ranges from 1/4 to ½ of a mile. However, distances exceeding 1/2 a mile, especially those greater than 2 miles, exhibit the highest negative impact on the rate of active transportation to school. However, only around 50% of CVESD students who lived within less than half a mile from the school chose to walk as their primary mode of transportation. Therefore, targeted strategies are required to encourage the remaining 50% to walk to school.

Mediating factors exert a significant influence on students' decisions regarding active transportation to and from school, with parental perceptions playing a crucial role. Our dataset indicates that 60% of students expressed a willingness to walk or cycle to school, seeking permission from their parents to do so. However, it is notable that only half of this group was observed to actively walk to school, suggesting that parents may have various underlying reasons for their decisions. Results from this study show that parental perceptions of the environment have a strong impact on transportation patterns to school, with concerns about crime rate and speed of traffic along the route as the most significant barriers to active transportation to school. Although factors such as the convenience of driving, after-school programs, safety of intersections and crossings, and traffic along the

route display relatively minor influences, they contribute slightly to the negative impact on the ATS rate. The presence of a crossing guard was also found to be a modest positive factor, often convincing more parents to allow their children to opt for active transportation. However, factors that one might anticipate influencing the decision, like concerns over time, the need for adult supervision during the walk, concerns regarding sidewalks and pathways, and weather conditions, surprisingly did not have even a minor impact on parents' choices.

Moreover, our study underscores the importance of schools' encouragement in promoting active transportation. Although the SRTS program period was not found to be a significant factor in increasing students' willingness to walk to school, low levels of ATS encouragement were found to adversely influence walking trends among students. School-based interventions and educational programs that foster positive attitudes towards active transportation can be instrumental in promoting sustainable and health-conscious travel choices among students. Additionally, student perceptions of ATS as uninteresting or boring are associated with reduced likelihoods of walking to school. This highlights the importance of creating engaging and enjoyable mechanisms to encourage students.

Furthermore, our study indicates that the possibility of opting for active transportation appears to be higher among male students compared to their female counterparts. Understanding the underlying reasons for this gender disparity can help inform gender-specific interventions to promote active transportation among female students.

Lastly, a few study limitations should be acknowledged. First, the low number of students riding a bike in our sample led us to exclude bicycle transportation from our analyses, potentially overlooking important factors influencing active transportation trends. Future studies with larger and more diverse datasets, including a sufficient representation of bicycling students, could provide a more comprehensive understanding of the determinants of transportation choices to school. Furthermore, our study focused on individual-level factors, and we did not analyze the influence of the Chula Vista Elementary School District's built environment on walking trends. Exploring the impact of the built environment, including infrastructure, sidewalk availability, and traffic safety measures, could provide valuable insights into how the surroundings affect active transportation behaviors. Addressing these limitations in future research will further enhance our understanding and inform targeted interventions aimed at fostering sustainable and health-conscious travel choices among students.

6. Conclusions

Leveraging the capabilities of various machine learning models, namely support vector machines (SVM), logistic regression (LR), and random forest (RF), this research utilizes SHAP (SHapley Additive exPlanations) values with the primary objective of understanding the determinants influencing transportation choices among students in the CVESD. SHAP offers a consistent and unified measure to interpret machine learning model outputs, aiding an in-depth, intuitive understanding of model decisions. Using SHAP in this study ensured an unbiased, accurate ranking of factors. The integration of advanced machine learning techniques in conjunction with SHAP contributed to a better understanding of the dynamics of student transportation. By identifying and highlighting barriers to ATS, this study provides policymakers with critical insights essential for crafting comprehensive, sustainable, and inclusive transport strategies. As cities and educational institutions globally strive for sustainable and inclusive transportation policies, understanding these nuanced factors becomes paramount.

A significant observation from the SHAP value rankings is its consistency with prior studies which emphasize the significant role of home-to-school distance in shaping students' transportation decisions. The data revealed that distances greater than 1/2 a mile, and especially those beyond 2 miles, had a profound negative influence on the likelihood of students choosing active transportation. Proximity to school continues to stand out as a primary determinant in students' transportation choices. Additionally, mediating factors,

particularly parental concerns regarding crime rates and speed of traffic along the route, emerged as critical influencers affecting parents' willingness to allow their children to engage in active transportation modes.

This study also underscores the importance of schools' encouragement and interventions in promoting active transportation among students. Evidently, the proactive involvement and policies of a school can significantly influence students' propensity to adopt active commuting methods. Interestingly, while the specific period of the Safe Routes to School (SRTS) program did not have a marked impact on students' choices, the overall encouragement from the schools had a pronounced effect. The study also found that even though around 60% of the students showed a willingness to walk, only half of them were actively doing so, revealing a potential gap between student intent and actual behavior. Gender disparities were also noted, with males more inclined to choose active transportation than females. Such gender-based differences prompt a need to further investigate the barriers faced specifically by female students and to tailor strategies accordingly. Furthermore, the performance metrics of our machine learning models showed promising results; both SVM and RF models demonstrated a prediction accuracy of 80%, outperforming the LR model. Such levels of accuracy suggest that the models can reliably predict transportation choices based on the given features.

We recommend that future research endeavors to incorporate larger and more diverse datasets and explore the influence of the built environment on walking behaviors. Such efforts hold the potential to yield valuable insights for targeted interventions aimed at promoting sustainable and health-conscious travel choices among students in the CVESD and similar settings.

Author Contributions: Conceptualization, B.E., A.J., S.G.M., G.F. and M.-H.T.; methodology, B.E., A.J. and S.G.M.; validation, B.E. and A.J.; formal analysis, B.E. and A.J.; resources, B.E., A.J. and G.F.; data curation, B.E., writing, B.E., A.J. and S.G.M.; visualization, B.E.; funding acquisition, G.F., A.J., S.G.M. and M.-H.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data are available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Poitras, V.J.; Gray, C.E.; Borghese, M.M.; Carson, V.; Chaput, J.-P.; Janssen, I.; Katzmarzyk, P.T.; Pate, R.R.; Connor Gorber, S.; Kho, M.E.; et al. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. *Appl. Physiol. Nutr. Metab.* **2016**, *41*, S197–S239. [CrossRef] [PubMed]
2. Janssen, I.; LeBlanc, A.G. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int. J. Behav. Nutr. Phys. Act.* **2010**, *7*, 40. [CrossRef]
3. Strong, W.B.; Malina, R.M.; Blimkie, C.J.; Daniels, S.R.; Dishman, R.K.; Gutin, B.; Hergenroeder, A.C.; Must, A.; Nixon, P.A.; Pivarnik, J.M.; et al. Evidence based physical activity for school-age youth. *J. Pediatr.* **2005**, *146*, 732–737. [CrossRef] [PubMed]
4. CDC. How Much Physical Activity Do Children Need? Centers for Disease Control and Prevention. 30 June 2023. Available online: https://www.cdc.gov/physicalactivity/basics/children/index.htm (accessed on 6 October 2023).
5. Guthold, R.; Stevens, G.A.; Riley, L.M.; Bull, F.C. Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1·6 million participants. *Lancet Child. Adolesc. Health* **2020**, *4*, 23–35. [CrossRef] [PubMed]
6. Hills, A.P.; Andersen, L.B.; Byrne, N.M. Physical activity and obesity in children. *Br. J. Sports Med.* **2011**, *45*, 866–870. [CrossRef] [PubMed]
7. Batista, M.B.; Romanzini, C.L.P.; Barbosa, C.C.L.; Blasquez Shigaki, G.; Romanzini, M.; Ronque, E.R.V. Participation in sports in childhood and adolescence and physical activity in adulthood: A systematic review. *J. Sports Sci.* **2019**, *37*, 2253–2262. [CrossRef]
8. Appleyard, B.S. Planning safe routes to school. *Planning* **2003**, *69*, 34–37.

9. Tudor-Locke, C.; Ainsworth, B.E.; Popkin, B.M. Active commuting to school: An overlooked source of childrens' physical activity? *Sports Med.* **2001**, *31*, 309–313. [CrossRef]

10. Faulkner, G.E.; Buliung, R.N.; Flora, P.K.; Fusco, C. Active school transport, physical activity levels and body weight of children and youth: A systematic review. *Prev. Med.* **2009**, *48*, 3–8. [CrossRef]

11. Davison, K.K.; Werder, J.L.; Lawson, C.T. Peer reviewed: Children's active commuting to school: Current knowledge and future directions. *Prev. Chronic Dis.* **2008**, 5.

12. SRTS G. SRTS Guide: The Decline of Walking and Bicycling. 2023. Available online: http://guide.saferoutesinfo.org/introduction/the_decline_of_walking_and_bicycling.cfm (accessed on 6 October 2023).

13. Babey, S.H.; Wolstein, J.; Diamant, A.L.; Bloom, A.; Goldstein, H. Overweight and Obesity among Children by California Cities-2010. 2012. Available online: https://escholarship.org/uc/item/7pm2s4k3 (accessed on 6 October 2023).

14. ATSDR. CDC/ATSDRCDC/ATSDRSVI Fact Sheet. Place Health. 2023. Available online: https://www.atsdr.cdc.gov/placeandhealth/svi/fact_sheet/fact_sheet.html (accessed on 6 October 2023).

15. Mennesson, M. Safe Routes to School: The Story Begins. Safe Routes Partnership. 2015. Available online: https://www.saferoutespartnership.org/blog/safe-routes-school-story-begins (accessed on 6 October 2023).

16. Coughlan, M.; Cronin, P.; Ryan, F. Survey research: Process and limitations. *Int. J. Ther. Rehabil.* **2009**, *16*, 9–15. [CrossRef]

17. Lizarazo, C.G.; Hall, T.; Tarko, A. Impact of the Safe Routes to School Program: Comparative analysis of infrastructure and noninfrastructure measures in Indiana. *J. Transp. Eng. Part. Syst.* **2021**, *147*, 04020151. [CrossRef]

18. Ragland, D.R.; Pande, S.; Bigham, J.; Cooper, J.F. Ten Years Later: Examining the Long-Term Impact of the California Safe Routes to School Program. 2014. Available online: https://escholarship.org/uc/item/8m59g6vx (accessed on 6 October 2023).

19. Orenstein, M.R.; Gutierrez, N.; Rice, T.M.; Cooper, J.F.; Ragland, D.R. Safe Routes to School Safety and Mobility Analysis. 2007. Available online: https://escholarship.org/uc/item/5455454c (accessed on 6 October 2023).

20. Boarnet, M.G.; Day, K.; Anderson, C.; McMillan, T.; Alfonzo, M. California's Safe Routes to School program: Impacts on walking, bicycling, and pedestrian safety. *J. Am. Plan. Assoc.* **2005**, *71*, 301–317. [CrossRef]

21. McDonald, N.C.; Steiner, R.L.; Lee, C.; Rhoulac Smith, T.; Zhu, X.; Yang, Y. Impact of the safe routes to school program on walking and bicycling. *J. Am. Plan. Assoc.* **2014**, *80*, 153–167. [CrossRef]

22. Buckley, A.; Lowry, M.B.; Brown, H.; Barton, B. Evaluating safe routes to school events that designate days for walking and bicycling. *Transp. Policy* **2013**, *30*, 294–300. [CrossRef]

23. McDonald, N.C.; Aalborg, A.E. Why parents drive children to school: Implications for safe routes to school programs. *J. Am. Plan. Assoc.* **2009**, *75*, 331–342. [CrossRef]

24. Dalton, M.A.; Longacre, M.R.; Drake, K.M.; Gibson, L.; Adachi-Mejia, A.M.; Swain, K.; Xie, H.; Owens, P.M. Built environment predictors of active travel to school among rural adolescents. *Am. J. Prev. Med.* **2011**, *40*, 312–319. [CrossRef]

25. McDonald, N.C. Active transportation to school: Trends among US school children, 1969–2001. *Am. J Prev. Med.* **2007**, *32*, 509–516. [CrossRef]

26. Mandic, S.; de la Barra, S.L.; Bengoechea, E.G.; Stevens, E.; Flaherty, C.; Moore, A.; Middlemiss, M.; Williams, J.; Skidmore, P. Personal, social and environmental correlates of active transport to school among adolescents in Otago, New Zealand. *J. Sci. Med. Sport.* **2015**, *18*, 432–437. [CrossRef]

27. De Meester, F.; Van Dyck, D.; De Bourdeaudhuij, I.; Deforche, B.; Cardon, G. Does the perception of neighborhood built environmental attributes influence active transport in adolescents? *Int. J. Behav. Nutr. Phys. Act.* **2013**, *10*, 38. [CrossRef]

28. Panter, J.R.; Jones, A.P.; Van Sluijs, E.M.; Griffin, S.J. Neighborhood, route, and school environments and children's active commuting. *Am. J. Prev. Med.* **2010**, *38*, 268–278. [CrossRef] [PubMed]

29. Hubsmith, D.A. Safe routes to school in the United States. *Child. Youth Environ.* **2006**, *16*, 168–190. [CrossRef]

30. Hopkins, D.; Mandic, S. Perceptions of cycling among high school students and their parents. *Int. J. Sustain. Transp.* **2017**, *11*, 342–356. [CrossRef]

31. Stewart, O. Findings from research on active transportation to school and implications for safe routes to school programs. *J. Plan. Lit.* **2011**, *26*, 127–150. [CrossRef]

32. Li, M.; Wang, Y.; Zhou, D. Effects of the built environment and sociodemographic characteristics on children's school travel. *Transp. Policy* **2023**, *134*, 191–202. [CrossRef]

33. Sadeghvaziri, E.; Javid, R.; Jeihani, M. Active Transportation for Underrepresented Populations in the United States: A Systematic Review of Literature. *Transp. Res. Rec.* **2023**, 03611981231197659. [CrossRef]

34. Rice, W.R. How We Got to School A Study of Travel Choices of Christchurch Primary School Pupils. 2008. Available online: https://ir.canterbury.ac.nz/items/51dc6b32-ad65-4550-8434-fd241f43d6f9 (accessed on 6 October 2023).

35. Huertas-Delgado, F.J.; Herrador-Colmenero, M.; Villa-González, E.; Aranda-Balboa, M.J.; Cáceres, M.V.; Mandic, S.; Chillón, P. Parental perceptions of barriers to active commuting to school in Spanish children and adolescents. *Eur. J. Public Health* **2017**, *27*, 416–421. [CrossRef] [PubMed]

36. Trang, N.H.; Hong, T.K.; Dibley, M.J. Active commuting to school among adolescents in Ho Chi Minh City, Vietnam: Change and predictors in a longitudinal study, 2004 to 2009. *Am. J. Prev. Med.* **2012**, *42*, 120–128. [CrossRef]

37. Chillón, P.; Ortega, F.B.; Ruiz, J.R.; Pérez, I.J.; Martín-Matillas, M.; Valtueña, J.; Gómez-Martínez, S.; Redondo, C.; Rey-López, J.P.; Castillo, M.J.; et al. Socio-economic factors and active commuting to school in urban Spanish adolescents: The AVENA study. *Eur. J. Public Health* **2009**, *19*, 470–476. [CrossRef]

38. Martin, S.L.; Lee, S.M.; Lowry, R. National prevalence and correlates of walking and bicycling to school. *Am. J. Prev. Med.* **2007**, *33*, 98–105. [CrossRef]

39. Pocock, T.; Moore, A.; Keall, M.; Mandic, S. Physical and spatial assessment of school neighbourhood built environments for active transport to school in adolescents from Dunedin (New Zealand). *Health Place* **2019**, *55*, 1–8. [CrossRef] [PubMed]

40. Chaufan, C.; Yeh, J.; Fox, P. The safe routes to school program in California: An update. *Am. J. Public Health* **2012**, *102*, e8–e11. [CrossRef] [PubMed]

41. Ahlport, K.N.; Linnan, L.; Vaughn, A.; Evenson, K.R.; Ward, D.S. Barriers to and facilitators of walking and bicycling to school: Formative results from the non-motorized travel study. *Health Educ. Behav.* **2008**, *35*, 221–244. [CrossRef] [PubMed]

42. Nelson, N.M.; Woods, C.B. Neighborhood perceptions and active commuting to school among adolescent boys and girls. *J. Phys. Act. Health* **2010**, *7*, 257–266. [CrossRef] [PubMed]

43. Woldeamanuel, M. Younger teens' mode choice for school trips: Do parents' attitudes toward safety and traffic conditions along the school route matter? *Int. J. Sustain. Transp.* **2016**, *10*, 147–155. [CrossRef]

44. Kerr, J.; Rosenberg, D.; Sallis, J.F.; Saelens, B.E.; Frank, L.D.; Conway, T.L. Active commuting to school: Associations with environment and parental concerns. *Med. Sci. Sports Exerc.* **2006**, *38*, 787–793. [CrossRef] [PubMed]

45. Hume, C.; Timperio, A.; Salmon, J.; Carver, A.; Giles-Corti, B.; Crawford, D. Walking and cycling to school: Predictors of increases among children and adolescents. *Am. J. Prev. Med.* **2009**, *36*, 195–200. [CrossRef]

46. Carver, A.; Timperio, A.; Hesketh, K.; Crawford, D. Are children and adolescents less active if parents restrict their physical activity and active transport due to perceived risk? *Soc. Sci. Med.* **2010**, *70*, 1799–1805. [CrossRef]

47. McDonald, N.C. Is there a gender gap in school travel? An examination of US children and adolescents. *J. Transp. Geogr.* **2012**, *20*, 80–86. [CrossRef]

48. Babey, S.H.; Hastert, T.A.; Huang, W.; Brown, E.R. Sociodemographic, family, and environmental factors associated with active commuting to school among US adolescents. *J. Public Health Policy* **2009**, *30*, S203–S220. [CrossRef]

49. Centers for Disease Control and Prevention (CDC). Barriers to children walking to or from school–United States, 2004. *MMWR Morb. Mortal. Wkly. Rep.* **2005**, *54*, 949–952.

50. Rahman, M.L.; Moore, A.; Smith, M.; Lieswyn, J.; Mandic, S. A conceptual framework for modelling safe walking and cycling routes to high schools. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3318. [CrossRef] [PubMed]

51. Ross, A.; Kurka, J.M. Predictors of Active Transportation Among Safe Routes to School Participants in Arizona: Impacts of Distance and Income. *J. Sch. Health* **2022**, *92*, 282–292. [CrossRef] [PubMed]

52. Chriqui, J.F.; Taber, D.R.; Slater, S.J.; Turner, L.; Lowrey, K.M.; Chaloupka, F.J. The impact of state safe routes to school-related laws on active travel to school policies and practices in US elementary schools. *Health Place* **2012**, *18*, 8–15. [CrossRef] [PubMed]

53. Elhenawy, M.; Rakha, H.A.; El-Shawarby, I. Enhanced modeling of driver stop-or-run actions at a yellow indication: Use of historical behavior and machine learning methods. *Transp. Res. Rec.* **2014**, *2423*, 24–34. [CrossRef]

54. Rasouli, S.; Timmermans, H.J. Using ensembles of decision trees to predict transport mode choice decisions: Effects on predictive success and uncertainty estimates. *Eur. J. Transp. Infrastruct. Res.* **2014**, *14*, 412–424. [CrossRef]

55. Ermagun, A.; Rashidi, T.H.; Lari, Z.A. Mode choice for school trips: Long-term planning and impact of modal specification on policy assessments. *Transp. Res. Rec.* **2015**, *2513*, 97–105. [CrossRef]

56. Cheng, L.; Chen, X.; De Vos, J.; Lai, X.; Witlox, F. Applying a random forest method approach to model travel mode choice behavior. *Travel. Behav. Soc.* **2019**, *14*, 1–10. [CrossRef]

57. Dave, S.M.; Raykundaliya, D.P.; Shah, S.N. Modeling trip attributes and feasibility study of co-ordinated bus for school trips of children. *Procedia Soc. Behav. Sci.* **2013**, *104*, 650–659. [CrossRef]

58. Assi, K.J.; Shafiullah, M.; Nahiduzzaman, K.M.; Mansoor, U. Travel-to-school mode choice modelling employing artificial intelligence techniques: A comparative study. *Sustainability* **2019**, *11*, 4484. [CrossRef]

59. Etaati, B.; Fernandez, G.; Mercado, A.; Jahangiri, A.; Tsou, M.-H.; Machiani, S.G. Evaluating the Safe Routes to School (SRTS) transportation program in socially vulnerable communities in San Diego County, California. *Safe Natl. UTC.* **2023**.

60. Agran, P. Walking and Biking to School: How to Keep Kids Safe. HealthyChildren.org. 2023. Available online: https://www.healthychildren.org/English/safety-prevention/on-the-go/Pages/Safety-On-The-Way-To-School.aspx (accessed on 8 October 2023).

61. Morrongiello, B.A.; Barton, B.K. Child pedestrian safety: Parental supervision, modeling behaviors, and beliefs about child pedestrian competence. *Accid. Anal. Prev.* **2009**, *41*, 1040–1046. [CrossRef] [PubMed]

62. Schieber, R.; Vegega, M. Education versus environmental countermeasures. *Inj. Prev.* **2002**, *8*, 10–11. [CrossRef] [PubMed]

63. Vatcheva, K.P.; Lee, M.; McCormick, J.B.; Rahbar, M.H. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiol. Sunnyvale Calif.* **2016**, *6*, 227. [CrossRef] [PubMed]

64. Tu, Y.K.; Clerehugh, V.; Gilthorpe, M.S. Collinearity in linear regression is a serious problem in oral health research. *Eur. J. Oral. Sci.* **2004**, *112*, 389–397. [CrossRef] [PubMed]

65. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* **1995**, *14*, 1137–1145.

66. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **2006**, *7*, 91. [CrossRef]

67. Kim, K.; Li, L. Modeling fault among bicyclists and drivers involved in collisions in Hawaii, 1986–1991. *Transp. Res. Rec.* **1996**, *1538*, 75–80. [CrossRef]

68. Mercier, C.R.; Shelley, M.C.; Rimkus, J.B.; Mercier, J.M. Age and gender as predictors of injury severity in head-on highway vehicular collisions. *Transp. Res. Rec.* **1997**, *1581*, 37–46. [CrossRef]

69. Hilakivi, I.; Veilahti, J.; Asplund, P.; Sinivuo, J.; Laitinen, L.; Koskenvuo, K. A sixteen-factor personality test for predicting automobile driving accidents of young drivers. *Accid. Anal. Prev.* **1989**, *21*, 413–418. [CrossRef]

70. James, J.L.; Kim, K.E. Restraint use by children involved in crashes in Hawaii, 1986–1991. *Transp. Res. Rec.* **1996**, *1560*, 8–12. [CrossRef]

71. Al-Ghamdi, A.S. Using logistic regression to estimate the influence of accident factors on accident severity. *Accid. Anal. Prev.* **2002**, *34*, 729–741. [CrossRef] [PubMed]

72. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.

73. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

74. Evans, J.; Waterson, B.; Hamilton, A. Forecasting road traffic conditions using a context-based random forest algorithm. *Transp. Plan. Technol.* **2019**, *42*, 554–572. [CrossRef]

75. Hamad, K.; Al-Ruzouq, R.; Zeiada, W.; Abu Dabous, S.; Khalil, M.A. Predicting incident duration using random forests. *Transp. Transp. Sci.* **2020**, *16*, 1269–1293. [CrossRef]

76. Xie, J.; Zhu, M. Maneuver-based driving behavior classification based on random forest. *IEEE Sens. Lett.* **2019**, *3*, 7002104. [CrossRef]

77. Yan, M.; Shen, Y. Traffic accident severity prediction based on random forest. *Sustainability* **2022**, *14*, 1729. [CrossRef]

78. Lu, Z.; Long, Z.; Xia, J.; An, C. A random forest model for travel mode identification based on mobile phone signaling data. *Sustainability* **2019**, *11*, 5950. [CrossRef]

79. Jahangiri, A.; Rakha, H. Developing a support vector machine (SVM) classifier for transportation mode identification by using mobile phone sensor data. In Proceedings of the Transportation Research Board 93rd Annual Meeting, Washington, DC, USA, 11–16 January 2014; Volume 14, p. 1442.

80. Jahangiri, A.; Rakha, H.A. Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2406–2417. [CrossRef]

81. Vanajakshi, L.; Rilett, L.R. Support vector machine technique for the short term prediction of travel time. In Proceedings of the 2007 IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, 13–15 June 2007; IEEE: New York, NY, USA, 2007; pp. 600–605.

82. Boswell, D. Introduction to support vector machines. *Dep. Comput. Sci. Eng. Univ. Calif. San. Diego.* **2002**, *11*.

83. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

84. Syarif, I.; Prugel-Bennett, A.; Wills, G. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *TELKOMNIKA Telecommun. Comput. Electron. Control.* **2016**, *14*, 1502–1509. [CrossRef]

85. Huang, C.L.; Wang, C.J. A GA-based feature selection and parameters optimizationfor support vector machines. *Expert. Syst. Appl.* **2006**, *31*, 231–240. [CrossRef]

86. Huang, J.; Ling, C.X. Constructing New and Better Evaluation Measures for Machine Learning. *IJCAI* **2007**, 859–864.

87. MacKay, D.J. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.

88. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 1.

89. Halimu, C.; Kasem, A.; Newaz, S.S. Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. In Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, Da Lat Viet Nam, Vietnam, 25–28 January 2019; pp. 1–6.

90. Cortes, C.; Mohri, M. Confidence intervals for the area under the ROC curve. *Adv. Neural Inf. Process Syst.* **2004**, *17*.

91. Lundberg, S.M.; Erion, G.G.; Lee, S.I. Consistent individualized feature attribution for tree ensembles. *arXiv* **2018**, arXiv:180203888.

92. Marcílio, W.E.; Eler, D.M. From explanations to feature selection: Assessing SHAP values as feature selection mechanism. In Proceedings of the 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Recife, Brazil, 7–10 November 2020; IEEE: New York, NY, USA, 2020; pp. 340–347.

93. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process Syst.* **2017**, *30*.

94. Tripathi, S.; Hemachandra, N.; Trivedi, P. Interpretable feature subset selection: A Shapley value based approach. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; IEEE: New York, NY, USA, 2020; pp. 5463–5472.

95. Ghaheri, P.; Nasiri, H.; Shateri, A.; Homafar, A. Diagnosis of parkinson's disease based on voice signals using SHAP and hard voting ensemble method. *Comput. Methods Biomech. Biomed. Engin.* **2023**, 1–17. [CrossRef]

96. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [CrossRef]