Trip purpose forecasting model for optimal energy management of electric vehicles supply equipment

1. Introduction

Electric vehicles (EVs) and plug-in hybrid EVs have experienced a surge in popularity within the automotive market, as indicated by recent statistics [1] and [2]. The increasing penetration of EVs raises concerns about uncoordinated charging behaviors, which can result in new load peaks on the aggregated power grid load curve. This, in turn, leads to various challenges, including power quality degradation [3], [4], and an increase in operational costs [5]. The integration of EVs into the distribution network poses a significant challenge for electric utilities, impacting the load profile, distribution system component capacity, voltage and frequency imbalances, excessive harmonic injection, power losses, and overall stability [6]. Consequently, scheduling EV charging becomes imperative.

The inherent unpredictability of user behaviors, particularly in scenarios where the power supply for EV supply equipment (EVSE) is limited, adds complexity to EV scheduling problems. Existing methods often overlook this limitation, assuming perfect knowledge of charging session parameters, such as stay duration and energy demand values, which is unrealistic in practice [6]. In a city setting, EVs can be categorized by use, including household EVs, passenger EVs, freight EVs, and those serving specific purposes such as postal or touristic needs [7]. While passenger and freight EVs may primarily charge at designated stations, a significant number of household EV owners lack private charging posts. Moreover, household EVs may require urgent charging during trips due to their less predictable schedules, making them the primary users of city-located quick charge stations. Understanding household EV behavior becomes crucial for assessing the charging load of charging stations [7].

The scheduling problem for EV charging behaviors introduces uncertainties related to start time, stay duration, and energy demand. These uncertainties cannot be entirely addressed through deterministic problem formulations. The real-time coordination of numerous EV charging behaviors is further complicated by the lack of a stochastic model capable of handling user behavior uncertainties [8]. The purpose of the trip emerges as a significant factor directly influencing stay duration. Given that household EVs share similar purposes with petrol vehicles, it is reasonable to assume that their behaviors align accordingly [9]. For instance, an EV entering an EVSE for work-related purposes may have a longer stay, allowing the operator to transfer energy at a lower rate, optimizing capacity utilization for EVs with shorter stay durations.

Machine learning approaches, predominantly utilizing decision trees [10-12] and based on variables computed for single activities, have shown promise. Deng and Ji reported accuracy rates ranging from 70% to 96% on a relatively homogeneous set of participants [11]. Lu et al. achieved accuracy rates between 60% and 73%, depending on the trip purpose [10] Liao et al. obtained good results (80% to 85% accuracy) using hierarchical, conditional random fields [13]. It is important to note that the reported classification accuracies differ in the level of detail for trip purposes, and the datasets vary in size and participant homogeneity.

Therefore, forecasting the purpose of a trip when an EV enters an EVSE can enable operators to make more optimal decisions for energy management. This project aims to develop a trip purpose forecasting model based on an origin-destination dataset, predicting the purpose of trips for enhanced energy management.

2. Problem Definition

The primary objective of this project is to address the challenges posed by uncoordinated charging behaviors of EVs through the development of a trip purpose forecasting model. The model aims to enhance energy management at EVSE locations.

2.1 Classification, Regression, or Clustering

The problem at hand involves a classification task. Given the inherent uncertainties in EV charging behaviors, particularly related to start time, stay duration, and energy demand, the model will classify the purpose of a trip when an EV enters an EVSE. The purpose classification will fall into distinct categories, such as work-related, household-related, or other specific purposes. This classification will assist EVSE operators in making optimal decisions for energy transfer rates based on the expected duration of the EV's stay.

2.2 Expected Input

The input features for the forecasting model will include variables that influence trip purposes. These variables may encompass the time of day, day of the week, historical charging behaviors, and potentially external factors such as weather conditions. Additionally, data related to the EV itself, such as its type (household, passenger, freight), battery level upon arrival, and the distance traveled since the last charge, will be considered. This multifaceted input dataset aims to capture the diverse factors influencing EV behaviors.

2.3 Expected Output

The model's output will be a classification of the purpose of the trip, assigning the EV to one of the predefined categories. This output is crucial for informing the charging strategy, allowing the EVSE operator to adjust energy transfer rates based on the expected stay duration associated with each trip purpose category.

2.1. Model Selection and Evaluation

To address this classification problem, three distinct machine learning models will be employed: linear models (Logistic Regression and LinearSVC), random forest, and neural network.

2.4.1 Linear Models (Logistic Regression, LinearSVM)

Logistic regression will be employed to establish a baseline model, exploring the linear relationships between input features and trip purposes. While logistic regression assumes a linear correlation, it provides a straightforward interpretation of feature importance and can serve as a benchmark for more complex models. LinearSVM will also be applied to potentially capture a more complicated decision boundary.

2.4.2 Random Forest

Random forest, an ensemble learning method, will be utilized to capture non-linear relationships and interactions among features. The flexibility of random forest in handling complex datasets makes it suitable for this problem, where the interplay of various factors influences the purpose of EV trips.

2.4.3 Neural Network

A neural network, specifically designed for classification tasks, will be implemented to capture intricate patterns in the data. The neural network's ability to learn hierarchical representations makes it well-suited for understanding the nuanced relationships among input features and trip purposes.

2.4.4 Evaluation Metrics

The performance of each model will be assessed using appropriate classification metrics such as accuracy, precision, recall, and F1 score. The chosen metrics will provide insights into the model's ability to correctly classify trip purposes and its overall effectiveness in enhancing energy management at quick charging stations. Cross-validation techniques will be employed to ensure robust model evaluation.

3. Dataset

The National Household Travel Survey (NHTS) stands out as a valuable resource for understanding transportation behaviors in North America, particularly in the United States [14]. With a substantial sample size of approximately 923,500 vehicle trips, the 2017 NHTS dataset provides a detailed snapshot of travel patterns and habits among individuals and households. It is conducted by the U.S. Department of Transportation's Federal Highway Administration, the 2017 NHTS dataset captures a wide array of variables related to transportation, including modes of transportation (car, public transit, walking, biking), trip durations, distances traveled, and purposes of travel (commuting, leisure, shopping, etc.). This wealth of information makes it a valuable tool for seeking into transportation dynamics for informed decision-making regarding infrastructure, urban planning, and transportation policies.

The dataset's notable feature is its extensive coverage, ensuring representation across diverse demographic and geographic characteristics. This inclusivity allows for statistically analyses and robust conclusions. Researchers can leverage the rich set of variables to gain insights into the complexities of transportation behavior in the United States, making informed decisions based on trends and policy impacts. The dataset categorizes the destination of each household EV trip into nine classes: Home, Shopping, Work, Social, Meals, Transport someone, School, Medical, and Others (see Fig. 1.) This categorization adds granularity to the analysis, offering a nuanced understanding of the diverse purposes behind household EV trips.

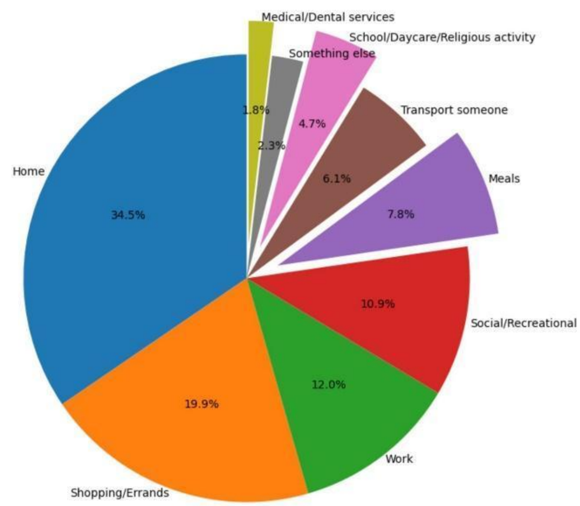


Fig. 1. The probability of trips based on trip purpose (NHTS 2017) [14]

In preparing the dataset for machine learning applications, several preprocessing steps may be considered. Imputation techniques could address missing data, ensuring a complete and representative dataset. Feature scaling may be applied, particularly for numerical variables, to prevent certain features from dominating the model. Categorical variables, like the destination classes, might undergo encoding for compatibility with machine learning algorithms. Data augmentation techniques could be explored to create synthetic variations, enhancing the model's generalization. Balancing classes ensures fair representation and prevents bias in the model's predictions.

The 2017 NHTS dataset, with its depth, breadth, and specific destination classes, forms the foundation for training and evaluating the trip purpose forecasting model. The dataset's characteristics are expected to contribute to the model's accuracy and generalizability, ultimately aiding in effective energy management at EV supply equipment locations.

4. Experiments

Data visualization

The first step in creating the random forest for prediction, or any machine learning solution, is often to plot the data and create some visualizations. Visualization helps you understand the characteristics of your data, identify patterns, and make informed decisions throughout the modeling process. The first step was to understand some of the features that were available in our large dataset. The dataset consists of ~923,000 entries of travel data, each with 115 features. In our case, these are encoded into field names by the Federal Highway Administration before becoming the columns of the dataset. For example, a feature called DWELTIME corresponds to the time at the destination in minutes. This created an extra step in understanding the data in that we had to use the NHTS codebook [15] to figure out what the columns meant. After this, the predicted category of WHYTRP1S, corresponding to the trip purpose summary, was chosen as our feature to predict. This was then plotted as seen in the figure below in order to get a better idea of the distribution of the predicted feature and understand its categories. Figure 2 (shown below) also helps to demonstrate why certain categories were much harder to predict as they had less data for the models to base off of. For example, the model had ~300,000 examples of a home trip but less than ~10,000 for a trip for medical or dental services.

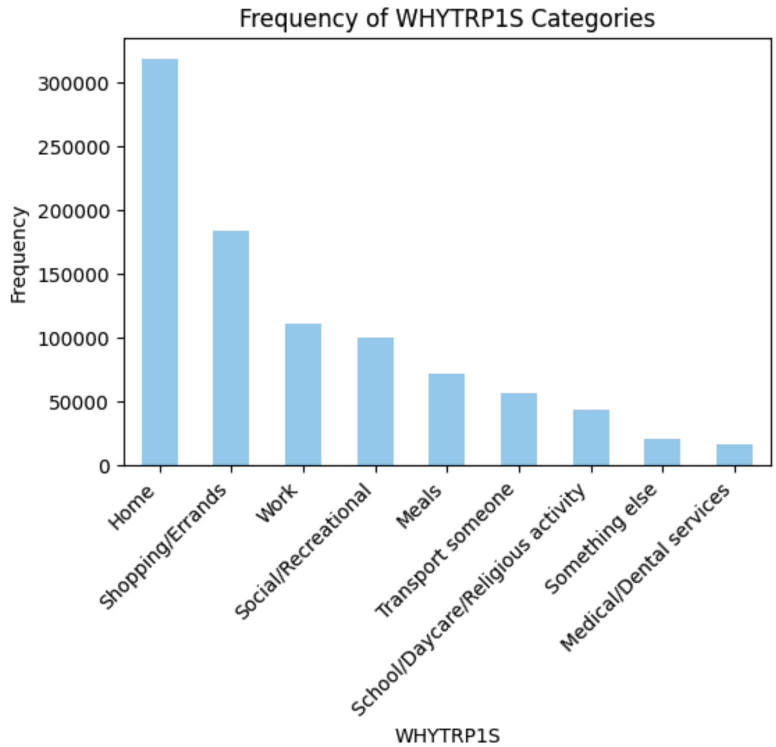
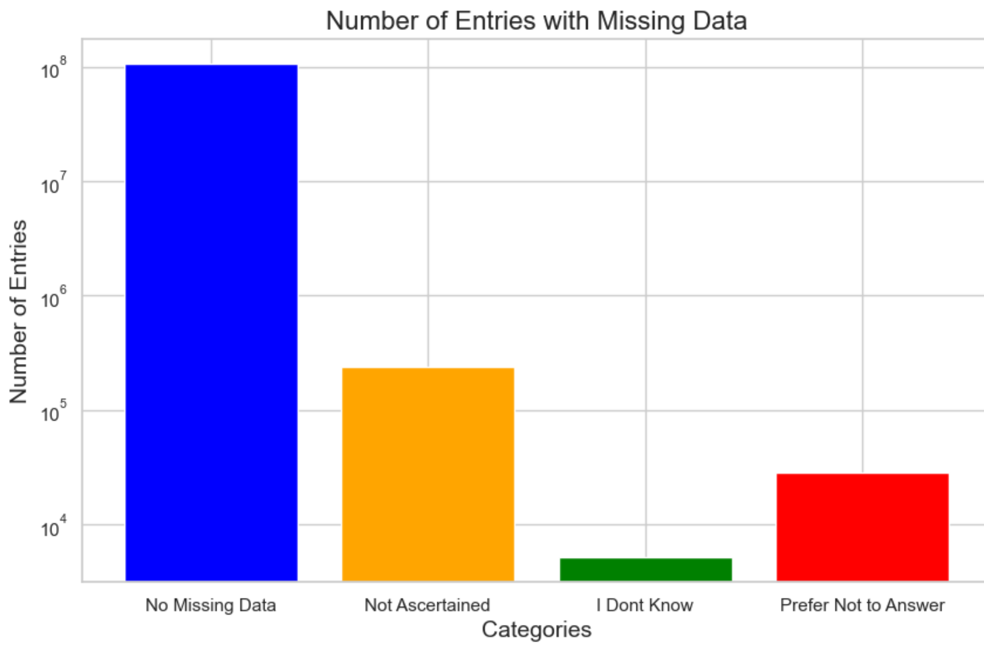


Fig. 2. Frequency of categories in the predicted feature of trip purpose

Preprocessing

After this was the preprocessing step, where certain columns were dropped in order to be able to use the data. For example, columns related directly to the trip purpose summary which are labeled in the data such as WHYTO, the trip destination purpose, had to be dropped since the purpose is to predict a very similar feature based on a real-world data scenario and not based on labels assigned by the dataset.

Next, we imputed missing data. While there were no NaN values in the dataset, for almost all features, the original survey provided the option to not respond to the question, and such responses were noted as special negative number codes in the final results. In addition, some data was “not ascertained” for unclear reasons and noted under a different code. A few other such codes exist in the dataset. None of these values provide any useful information about the feature, so we decided to impute them to retain as much information as possible. This strategy makes sense because we had so many samples available, which allows mean imputation to be reliable.



We then standardized the data using the StandardScaler available from sklearn. This involves subtracting the mean from each column and dividing by its standard deviation to ensure that each feature follows the standard normal distribution. Standardization is important when testing these models to ensure that the scale is consistent across all features and the model is not learning based on outlier magnitudes. It also helps with convergence and helps ensure that components such as the regularization terms do not affect certain features in disproportionate ways.

Finally, we one-hot encoded a few of the features that were categorical, such as home state. These steps allowed us to utilize as much of the data as possible and give our models the best chance to learn the correct relationship between its features and trip purpose.

Logistic Regression, LinearSVM

As a baseline test, we started with Logistic Regression, one of the simplest models for classification. We used the sklearn implementation, which comes with many choices for solver, loss, regularization, and other parameters. First, we found that the default solver lbfgs is slow to converge on our data due to size. As such, we randomly subsampled 100,000 data points and used a faster solver for large data (saga) to facilitate cross-validation. Another important feature of these models is the choice of C, the regularization parameter. It is ideal to search for the best value of C using cross-validation to achieve the best performance on testing data. We accomplished this using LogisticRegressionCV. The accuracy across categories on testing data is displayed below.

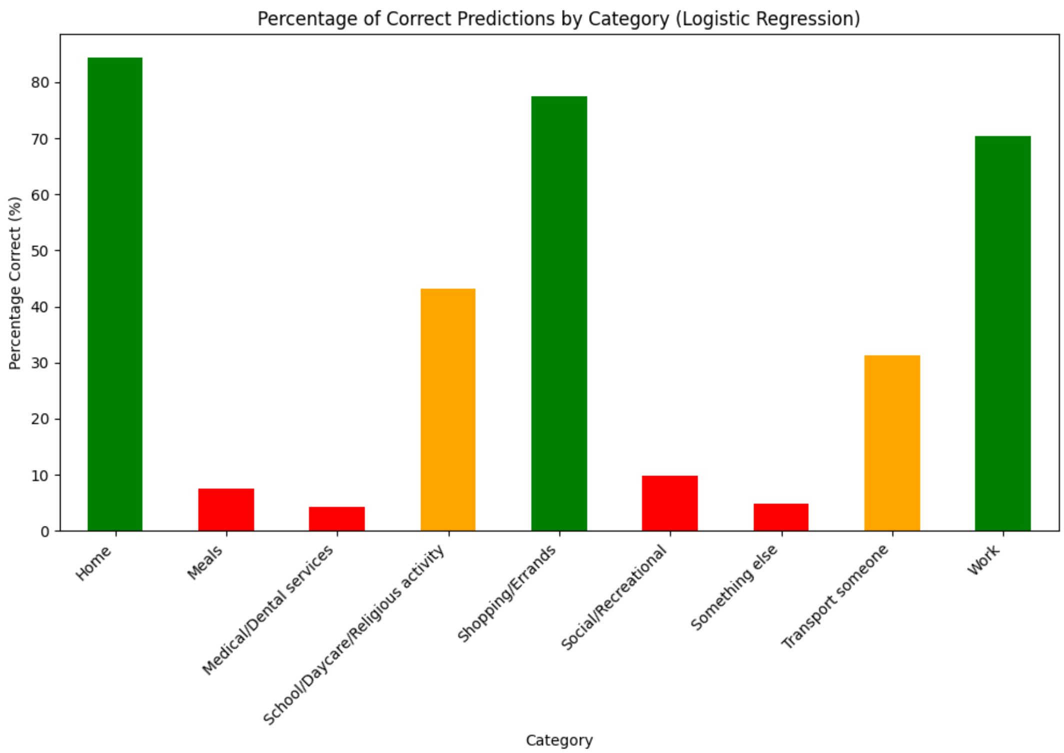
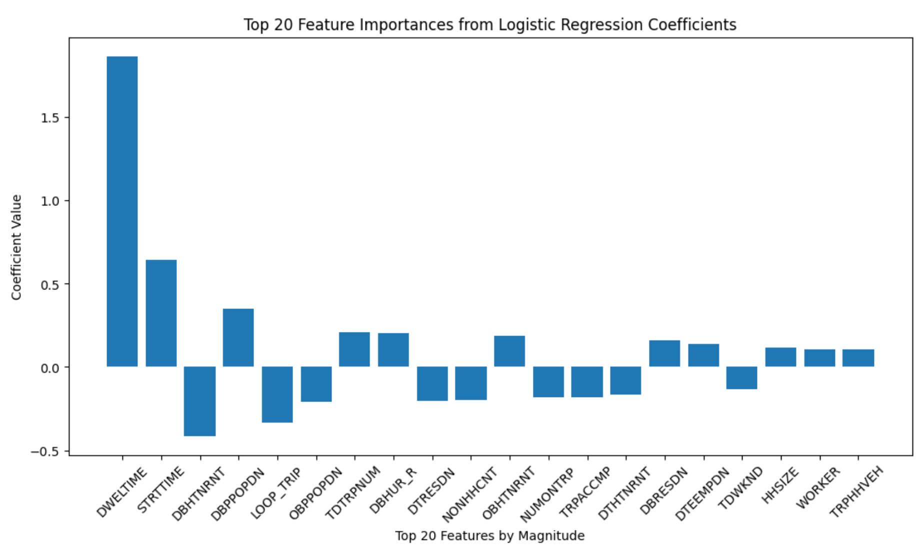
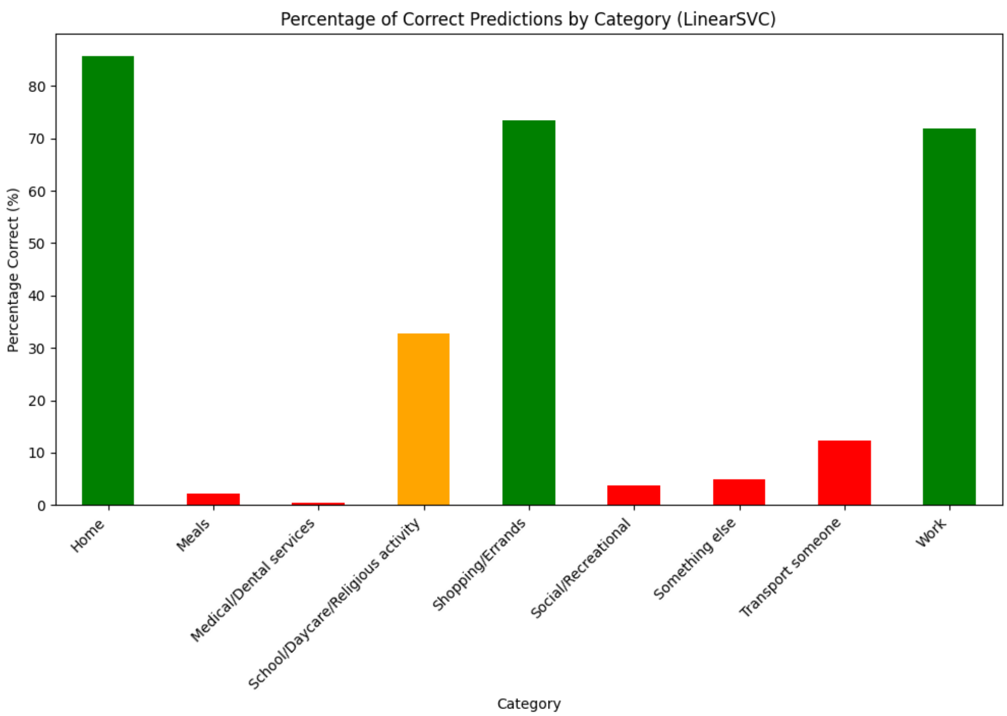


Fig. 3. Frequency of correct vs incorrect predictions for logistic regression model

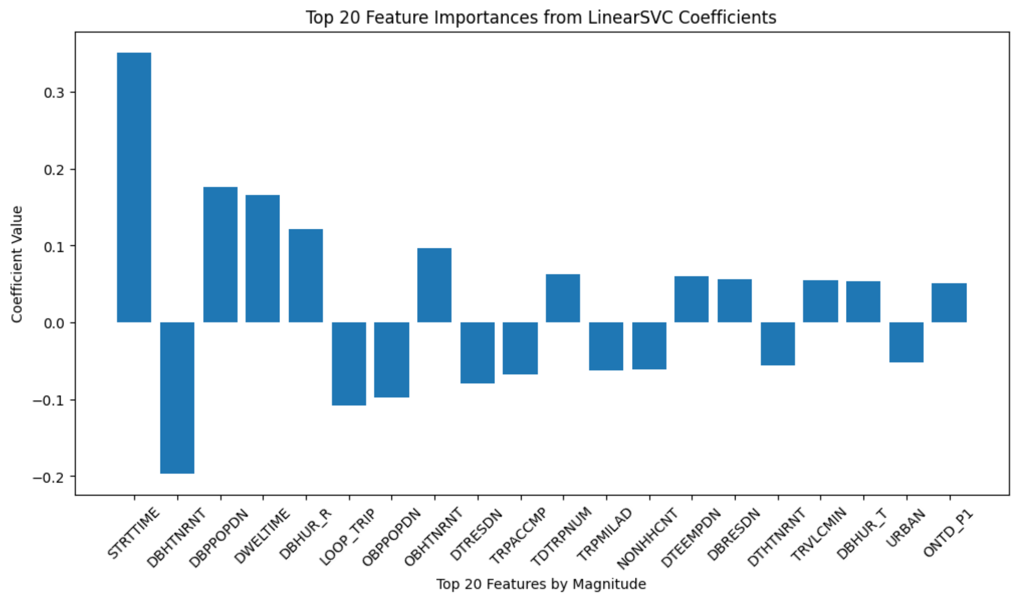
We noted that categories with low amounts of training data were much harder for the model to predict, as expected. The best logistic regression model obtained through cross validation had overall results of **59% correct predictions** on the test data. While tuning the regularization helped to some extent, the inability of the model to converge in a reasonable amount of time on larger datasets limits its power. In addition, it can only consider linear relationships between the data, which makes it harder to achieve higher accuracy in cases where some variables may be interacting with each other.



Following logistic regression, we moved on to a more advanced linear model using the Linear Support Vector Machine. These offer several advantages over logistic regression, especially when dealing with large datasets. LinearSVC excels in high-dimensional spaces, making it suitable for scenarios where the number of features is substantial. SVMs, in general, aim to find a hyperplane that maximally separates classes, making them effective in handling complex decision boundaries. In contrast, logistic regression models the probability of belonging to a particular class, which may not capture intricate relationships as effectively as SVMs in high-dimensional spaces. LinearSVC is robust to outliers and provides better performance when there is a clear margin of separation between classes. Additionally, SVMs are less sensitive to the multicollinearity of features compared to logistic regression.



Once again, categories with low amounts of training data were much harder for the model to predict, as expected. Surprisingly though, the best LinearSVC model obtained through cross validation had overall results of **56% correct predictions** on the test data, which is worse than the best Logistic Regression model. LinearSVC may underperform compared to logistic regression due to its sensitivity to outliers, which are quite probable in this data. The scalability issues for large data also prevented us from using more training data. Finally, we may have needed more computational power to tune hyperparameters given its sensitivity to hyperparameter choices. Once again, the feature importances are shown below, and at least somewhat match up with logistic regression:



Random Forest

Multiple random forests were created and tested over the course of the project to try and create the best predictive model for the dataset to predict the trip purpose. There were several steps in this process, including visualization, preprocessing, and model creation and testing on real data.

For the initial random forest, categorical data was dropped as well and most of the hyperparameters for the random forest were set to the defaults as given by the sklearn library. The results are as follows in figures 3 and 4.

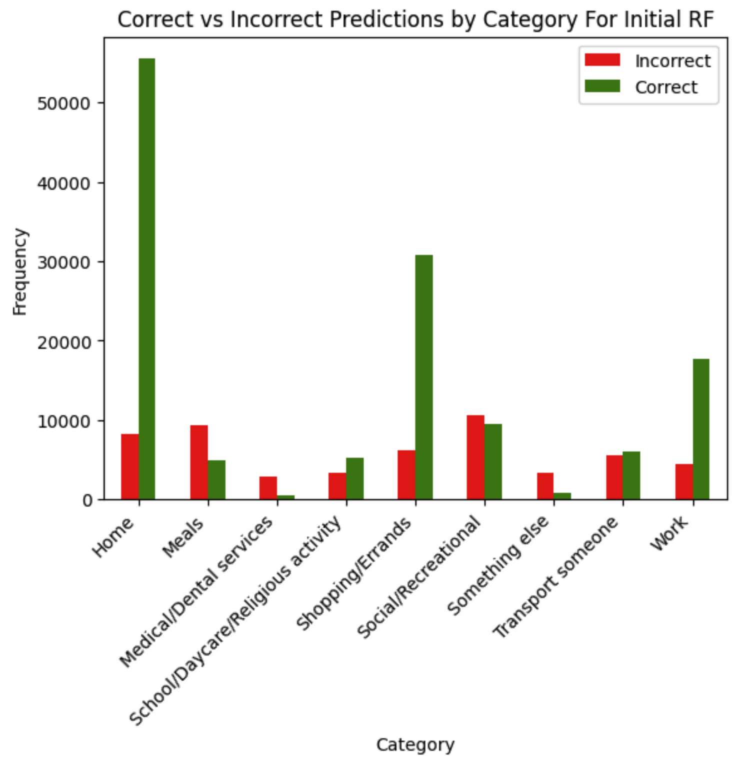


Fig. 3. Frequency of correct vs incorrect predictions for the initial random forest model

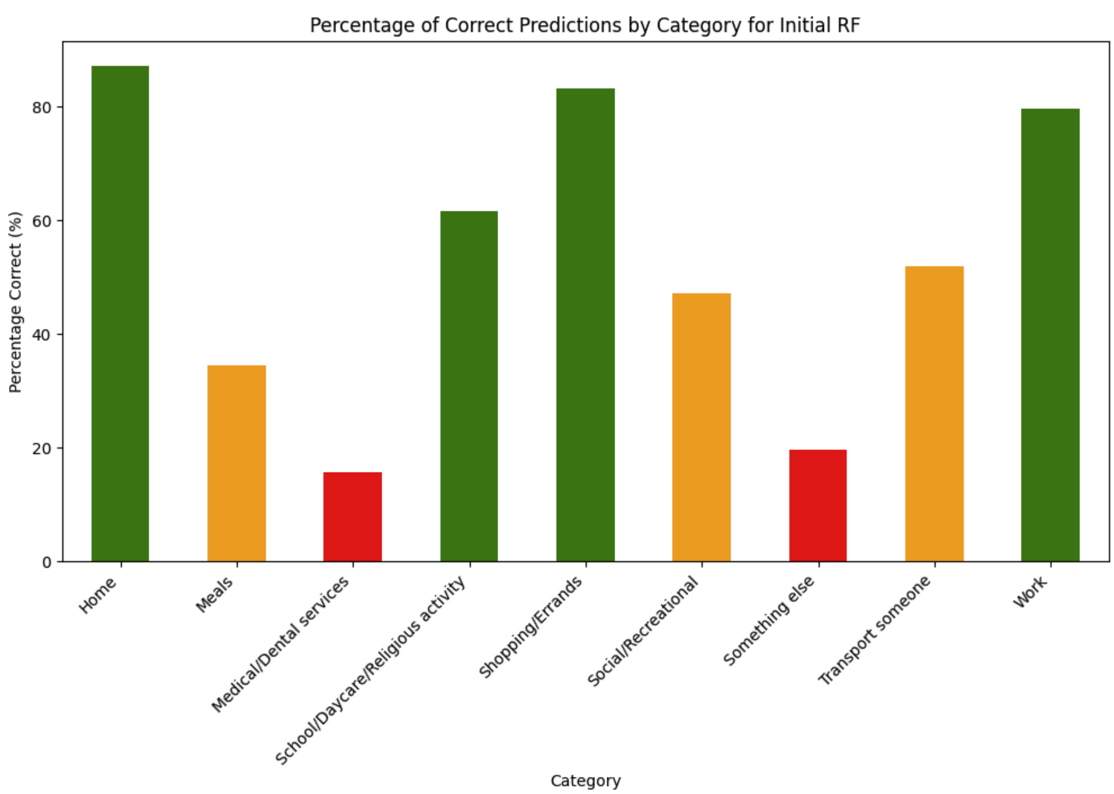


Fig. 4. Percentage of predictions correct for the initial random forest

After this, steps were taken to improve the processing of the data and to improve the model. One important step was the reduction of redundant variables, which were found by a correlation matrix for the most important features of the initial random forest, as shown below in figure 5.

	DWELTIME	STRTTIME	ENDTIME	TDTRPNUM	TRPMILAD	TRPMILES	R_AGE_IMP	R_AGE	WTTRDFIN	TDCASEID	WHYTRP1S
DWELTIME	1.00	-0.45	-0.44	-0.28	-0.00	-0.00	-0.13	-0.13	0.02	0.00	-0.14
STRTTIME	-0.45	1.00	0.97	0.53	-0.02	-0.02	-0.06	-0.06	0.01	-0.01	-0.08
ENDTIME	-0.44	0.97	1.00	0.51	0.02	0.02	-0.05	-0.05	0.01	-0.01	-0.08
TDTRPNUM	-0.28	0.53	0.51	1.00	-0.03	-0.03	0.06	0.06	-0.02	-0.02	-0.06
TRPMILAD	-0.00	-0.02	0.02	-0.03	1.00	1.00	0.00	0.00	-0.01	-0.00	0.05
TRPMILES	-0.00	-0.02	0.02	-0.03	1.00	1.00	0.00	0.00	-0.00	-0.00	0.05
R_AGE_IMP	-0.13	-0.06	-0.05	0.06	0.00	0.00	1.00	0.99	-0.17	0.01	0.05
R_AGE	-0.13	-0.06	-0.05	0.06	0.00	0.00	0.99	1.00	-0.17	0.01	0.05
WTTRDFIN	0.02	0.01	0.01	-0.02	-0.01	-0.00	-0.17	-0.17	1.00	-0.02	-0.01
TDCASEID	0.00	-0.01	-0.01	-0.02	-0.00	-0.00	0.01	0.01	-0.02	1.00	-0.01
WHYTRP1S	-0.14	-0.08	-0.08	-0.06	0.05	0.05	0.05	0.05	-0.01	-0.01	1.00

Fig. 5. Correlation matrix for the top features from the initial random forest

Categorical columns were also given much more value by choosing to encode them numerically instead of dropping the columns. For the random forest, this ended up being quite costly as the method chosen added over 100 columns to the dataset while one-hot encoding, bringing the total features for the processed data up to 222 (almost double the original count of 115).

After testing numerous random forests and validating the results, the best model obtained overall results of around **72% correct predictions** on the test data. Some major areas that improved the model were tuning the hyperparameter of the max number of features to be considered at each step in the tree and allowing it to be any of the now 222 features. This ran the risk of overfitting, but due to the ensemble's natural resistance to overfitting, did not become an issue. The results for this final model can be seen in the figures below.

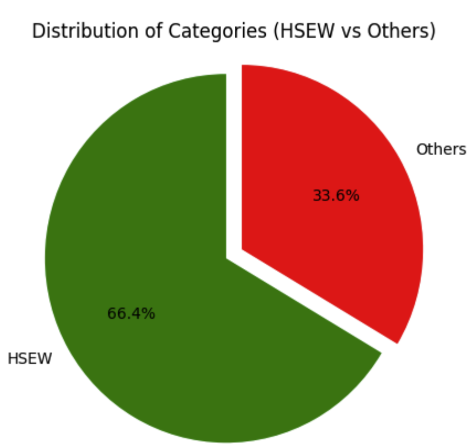


Fig. 6 and 7. Distribution of categories and correctness differences. These highlight that the model predicts with over 80% accuracy on the 3 majority trip destinations of home, shopping/errands, and work (HSEW).

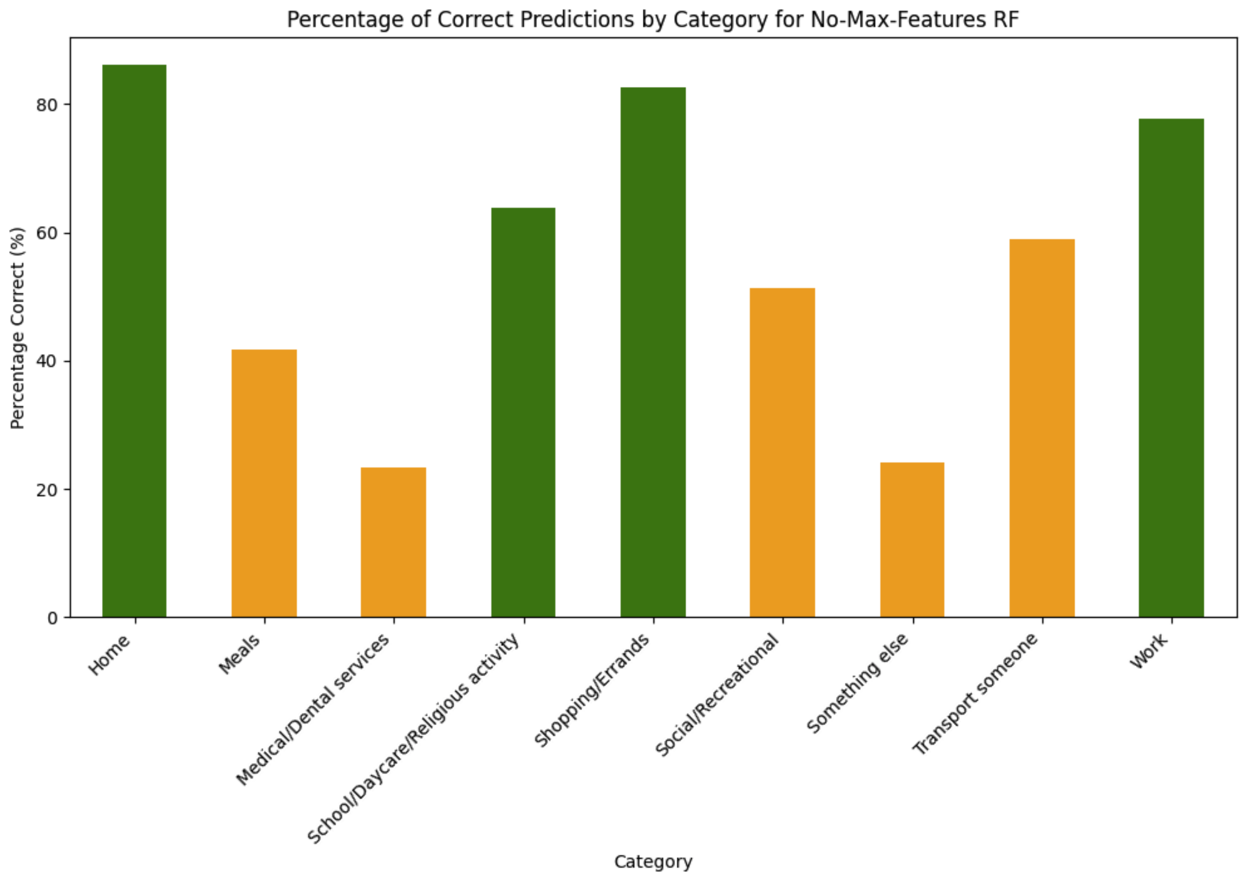
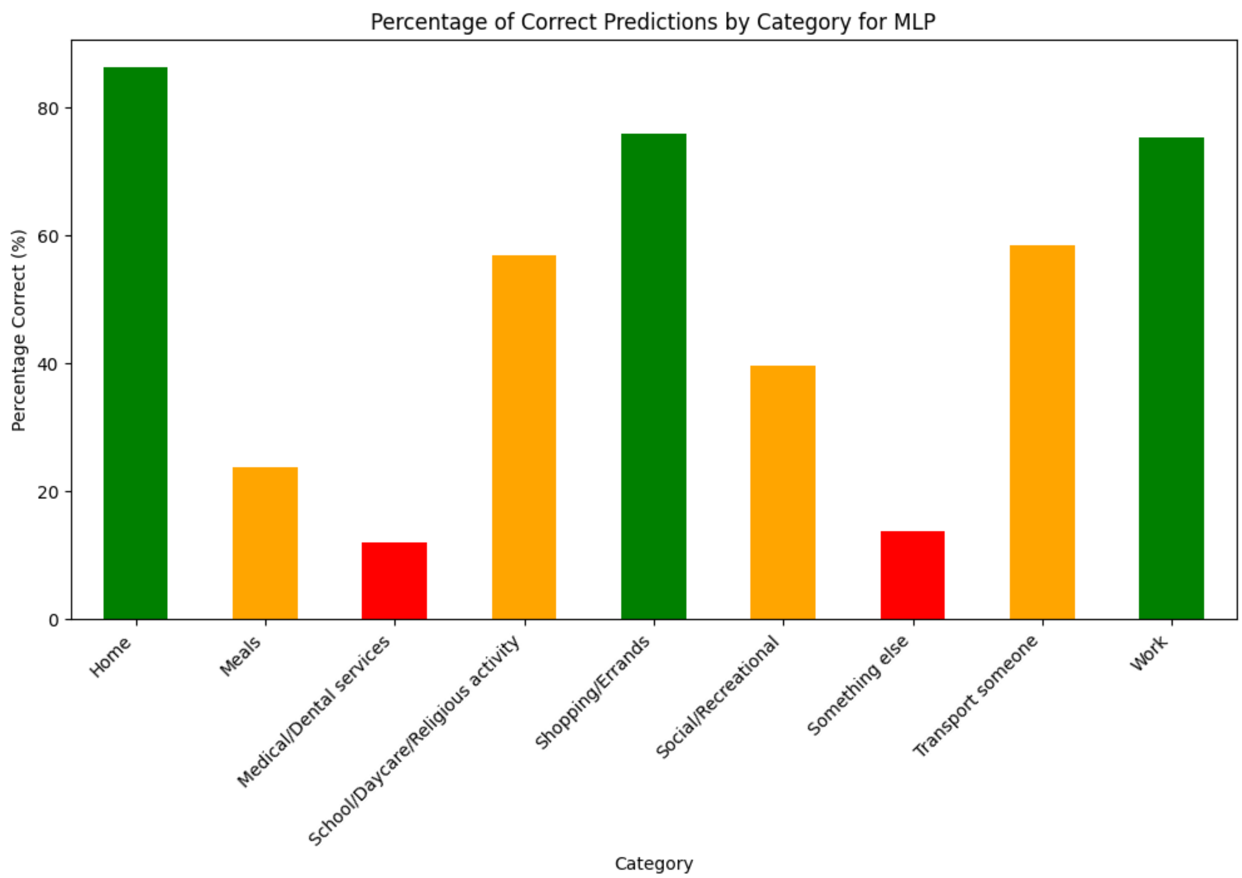


Fig. 8. Percentage of predictions correct for the final random forest

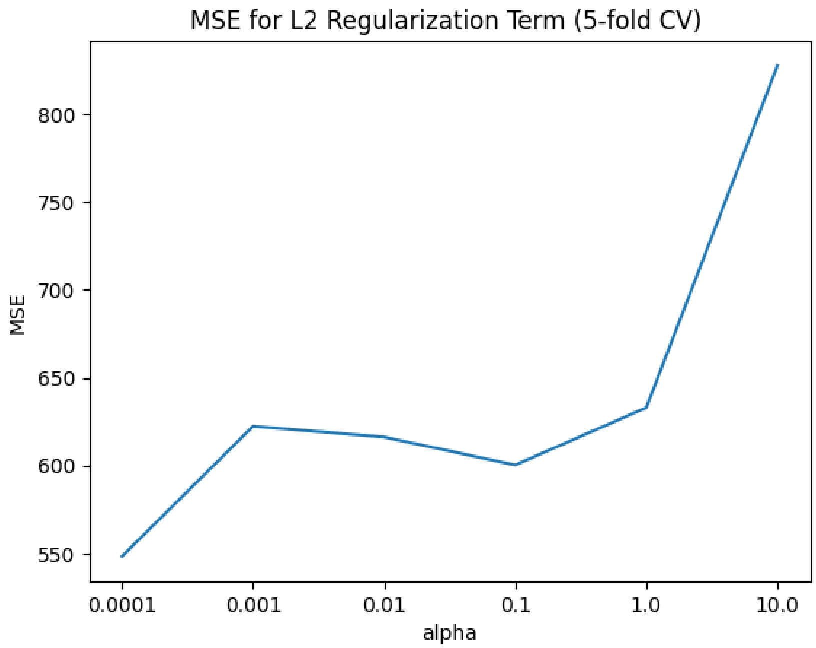
As shown in the figures above, the final random forest had a 72% accuracy. It was very good at certain categories of trip purpose, in particular it was incredibly strong at predicting the majority categories of HSEW (Home, Shopping/Errands, Work). HSEW makes up a majority of the dataset, covering around 66% of all trips made. On this 66% slice, the model was over 80% accurate at predicting the trip purpose. Other categories were weaker, but as shown by figure 8, all were improved significantly by improvement of the random forest when compared to the original results from figure 4.

Multi-layer Perceptron (Neural Network)

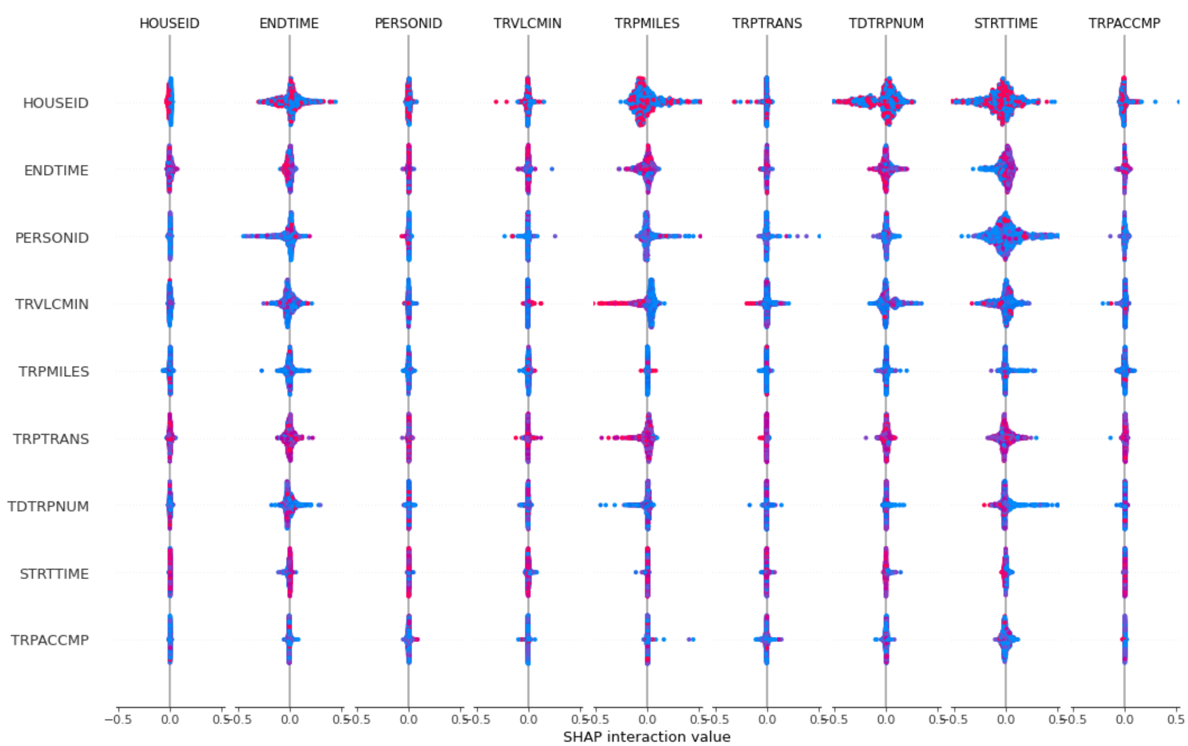
The last model we considered was a neural network. Specifically, we utilized the multi-layer perceptron classifier from the sklearn library. After implementing the preprocessing steps mentioned above, a neural network was fit to the training data. In the initial tests, all hyperparameters were kept as the default options (e.g. 1 hidden layer with 100 neurons, 0.0001 L2 regularization, etc.). This model achieved an **overall accuracy of 67%**. The results for each prediction category can be seen below.



Clearly, the strength of highly nonlinear models such as neural networks lies in the fact that its hyperparameters can be finely tuned to produce even better results. The ideal scenario would be to perform a grid search over all reasonable values for each parameter. However, due to both lack of computational resources and time (~ 1 million samples and 167 features after preprocessing), this study was restricted to a single hyperparameter deemed important for feature weighting. Namely, L2 regularization. With such variation in features even after standardizing, it is important to consider the impact this term has on model fitting and testing accuracy. Therefore, 5-fold cross validation was performed for alpha terms within the log space of 0.0001 - 10. As seen, however, the alpha that performed the best was actually the default value in sklearn, 0.0001:



Finally, a further analysis of the feature impacts on predicting trip purpose was performed. The results are shown in the following SHAP figure. SHAP is an explanation method used to visualize feature importances and can be used on ML models (especially black boxes such as neural networks) to provide better interpretability for the results.



This plot shows 9 example features (the first 9 in the dataframe) and their corresponding importances. However, it differs from standard feature importance plots because it shows the interaction shap value for each pair, something we did not consider previously in the logistic regression and random forest analyses. If a feature has little variation along its respective column, this signifies that the interaction of those features does not have much effect in predicting trip purpose. Further, the intersection of two features shows their relationship with each other. This plot intuitively makes sense, as it shows that features such as House ID and Person ID have almost no variation/importance in predicting trip purpose. They are simply used to label individual houses and people. Meanwhile, features such as start time, end time, and trip miles vary greatly. These are much more likely to influence the purpose of travel. If one looks at the larger clusters, however, these interactions indicate that certain people/houses could be associated with specific start/end times as well, which could indicate that some people may be following a regular schedule. A common example of this is an individual with a daily commute to work, who might have a consistent start time, end time, and miles traveled.

Despite the great variation that we see in those interactions, the plot is still relatively uninformative because it suggests that for specific samples, the interaction features can help, while for other samples, it might actually negatively affect the prediction. We would expect the most helpful interactions to lie on the 0 line or above, but there is not any strong evidence of this for any of the features. We chose to display these features after looking at several other interaction combinations and not finding any other significant results.

5. Future Work

Building on the current project's foundation, the next steps involve developing a comprehensive predictive framework for estimating the stay duration of EVs at EVSE locations. Leveraging the insights gained from predicting trip purposes using the NHTS dataset, future work entails integrating features from various datasets, including electrical utilities, and user feedback. Advanced machine learning techniques, such as reinforcement learning and time-series analysis, will be explored to capture temporal aspects and optimize dynamic charging infrastructure. The scalability and generalizability of the model will be a focus, ensuring its applicability across diverse geographic regions. The envisioned future work aims to contribute to a smarter and more responsive charging infrastructure, ultimately fostering the widespread adoption of EVs.

6. Limitations

The current model and report is simply a test of the predictive capabilities of a model trained in trip purpose predictions and how that prediction could be useful to the EV industry. However, applying this to the real-world would have many issues. One major problem is the reduction in real-world data and knowledge at the time of EV charging in order to predict the trip purpose and therefore optimize the charging of EVs at a charging station. The trip data used here is taken after the trip and has more available data, such as the start and end time of the trip, which is then used in our models for the prediction. To apply this to a real-world scenario would be much more difficult since we would have other, different features to work off of. This data could come from the EV itself, the user's history, or other sources. Despite these limitations, such a model could be possible and, if implemented correctly, could provide major benefits to the rising industry of the electrification of transportation.

GitHub:

https://github.com/prateekanand2/Travel-Purpose-Prediction/tree/main

References

[1] Ou, Shiqi, et al. "Light-duty plug-in electric vehicles in China: An overview on the market and its comparisons to the United States." Renewable and Sustainable Energy Reviews 112 (2019): 747-761.

[2] Wolinetz, Michael, and Jonn Axsen. "How policy can build the plug-in electric vehicle market: Insights from the REspondent-based Preference And Constraints (REPAC) model." Technological Forecasting and Social Change 117 (2017): 238-250.

[3] R.-C. Leou, C.-L. Su, and C.-N. Lu, "Stochastic analyses of electric vehicle charging impacts on distribution network," IEEE Trans. Power Syst., vol. 29, no. 3, pp. 1055–1063, May 2014.

[4] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," IEEE Trans. Power Syst., vol. 25, no. 1, pp. 371–380, Feb. 2010.

[5] S. Bashash and H. K. Fathy, "Cost-optimal charging of plug-in hybrid electric vehicles under time-varying electricity price signals," IEEE Trans. Intell. Transp. Syst., vol. 15, no. 5, pp. 1958–1968, Oct. 2014. [6] Das, Himadry Shekhar, et al. "Electric vehicles standards, charging infrastructure, and impact on grid integration: A technological review." Renewable and Sustainable Energy Reviews 120 (2020): 109618.

[7] Liu, Zeyu, et al. "Load forecasting model and day-ahead operation strategy for city-located EV quick charge stations." (2019): 237-6.

[8] Wang, Bin, et al. "Predictive scheduling framework for electric vehicles with uncertainties of user behaviors." *IEEE Internet of Things Journal* 4.1 (2016): 52-63.

[9] Zhou, Yun, Zheng Yan, and Naihu Li. "A novel state of charge feedback strategy in wind power smoothing based on short-term forecast and scenario analysis." IEEE Transactions on Sustainable Energy 8.2 (2016): 870-879.

[10] Lu, Y., S. Zhu, and L. Zhang. A Machine Learning Approach to Trip Purpose Imputation in GPS-Based Travel Surveys. Presented at 4th Conference on Innovations in Travel Modeling, Tampa, Fla., 2012.

[11] Deng, Z., and M. Ji. Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. Presented at 7th International Conference on Traffic and Transportation Studies, Kunming, China, 2010.

[12] Griffin, T., and Y. Huang. A Decision Tree Classification Model to Automate Trip Purpose Derivation. Presented at 18th International Conference on Computer Applications in Industry and Engineering, Honolulu, Hawaii, 2005.

[13] Liao, L., D. Fox, and H. Kautz. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields. International Journal of Robotics Research, Vol. 26, No. 1, 2007, pp. 119–134.

[14] https://nhts.ornl.gov/downloads

[15] https://nhts.ornl.gov/tables09/CodebookBrowser.aspx