Impact of Electric Vehicle Adoption on Electricity Distribution and Grid Demand

Nina Vuide, Eric Davidson

Department of Business, Florida Atlantic University

ISM 6136: Data Mining and Predictive Analytics

Prof. Magno Queiroz

November 25th 2024

Table of Contents

Executive Summary and Key Findings…………………………………………………………03

Description of Business Opportunities…………………………………………………………..03

Data for Analysis………………………………………………………………………………...05

Dataset Review EV Intelligent Port Logistics…………………………………………………05

Discuss and justify model assessment and selection…………………………………..07

Dataset Review Electric Vehicle Charging Patterns……………………………………………..10

Dataset Review…………………………………………………………………………12

Conclusions from Model Assessment…………………………………………………………..13

Recommendations for Management…………………………………………………………….14

Works Cited .................................................................................................................................. 17

Impact of Electric Vehicle Adoption on Electricity Distribution and Grid Demand

1. Executive Summary and Key Findings

The rapid adoption of electric vehicles (EVs) globally poses significant implications for electricity distribution and grid demand. This paper explores the relationship between EV adoption and its effects on electricity grid demand, focusing on charging station behavior and utility rate structures. Using datasets from the IEA Global EV Data (2024), charging station data, and utility rate information for both investor-owned (IOU) and non-investor-owned (non-IOU) utilities, this study investigates charging behavior patterns, utility rate impacts, and potential grid load concerns. Key findings include:

1. Most EV charging sessions result in low kWh consumption, indicating frequent but short duration charging.
2. Utility rate structures have a potential influence on charging behavior; however, merging utility data with station data revealed challenges due to unmatched zip codes.
3. A predictive model was attempted using regression approaches but faced data integration limitations that need resolution to enable more accurate forecasting.

2. Description of Business Opportunities

Business Problem and Opportunity

The adoption of EVs presents a critical opportunity for electric utilities, grid operators, and policymakers to optimize electricity distribution and balance grid load. As EV penetration grows, understanding charging behaviors and their impact on grid demand becomes essential for strategic infrastructure planning and rate structuring. The purpose of this data mining task is to analyze and predict the impact of EV charging on electricity grid demand using relevant data sources.

One area to consider would be its increased demand for electricity, as more EVs are sold items such as charging loads are a resource that are paramount and will need to be considered. Just as gas stations are abundant, one area that we could hope to see improvement is the rate at which vehicles can refuel. Currently most petrol vehicles can be refiled within 5 minutes where contrarily EVs can take as little as 30 mins (depending on expected travel distance) to hours depending on charging speed.

Revenue Implications for Utilities: With more people charging EVs, utilities have the potential to experience changes in revenue streams, especially if EVs reduce overall energy consumption at certain times or if charging costs are shared between utilities, third-party operators, and consumers. Utilities may partner with private charging infrastructure providers to create synergies in grid management and incentivize off-peak charging through lower rates or tiered pricing models. This could tie into incentives either subsidized by the government (local or federal) or by the utility companies to ease loads at peak times.

Some areas for concern could be the adverse effect placing a strain on Low-Income communities. In some cases, the costs associated with EVs (purchasing the vehicle, home charging equipment, and energy costs) may disproportionately affect low-income households. This could lead to an imbalanced across different regions, as in there may be an uneven distribution of charging stations, leading to "charging deserts" in some areas, particularly in rural or underdeveloped locations.

3. Data for Analyses

Description of the Datasets

1. Non-IOU and IOU Utility Data (2022):

* Description: Provides information on utility rates, service types, and ownership by zip code.

Initial Impressions of the Data

* The IEA EV Data provided extensive coverage of EV trends and shares.
* The charging station data contained detailed session-level attributes but had minimal missing data (in `distance`).
* The utility datasets included comprehensive rate data by zip code for non-IOU and IOU utilities, but these did not successfully merge with charging data due to unmatched zip codes.

2. EV Intelligent Port Logistics

This dataset provides records of electric vehicle (EV) charging loads, specifically collected from logistics and intelligent port operations in the greater Dallas, Texas area. It provides a comprehensive set of features that allow for deep analysis and predictive modeling in EV fleet management, energy load forecasting, and smart grid integration. The hope for this dataset is to have a better understanding of EV usage patterns in these sectors and their impact on the grid and communities they support. https://www.kaggle.com/datasets/datasetengineer/ev-intelligentport-logistics

Initial impressions of this data set appear to be promising, primarily looking at one attribute Grid\_Demand\_MW. This value reports the overall grid demand (in megawatts) at the time, important for understanding the relationship between EV charging and grid load balancing. This along with EV usage patterns, charging preferences and existing charging incentives can all play a pivotal role in shaping users’ habits to minimize load strain on existing resources.

When it comes to assessing the raw data, this dataset was decent in terms of competition but there were a few instances of missing values. In attempt to keep the data relatively normalized we had chosen to replace these missing values with the average for the attribute. When it came to prepping this data for regression, we had to convert these polynomial values to numeric/integers to determine the strength of the relationships. We felt it was best to keep values such as temperature, and rain fall in Celsius and millimeters respectively as these conversions would not affect any outcome and with majority of the world using metric it would be more useful kept as

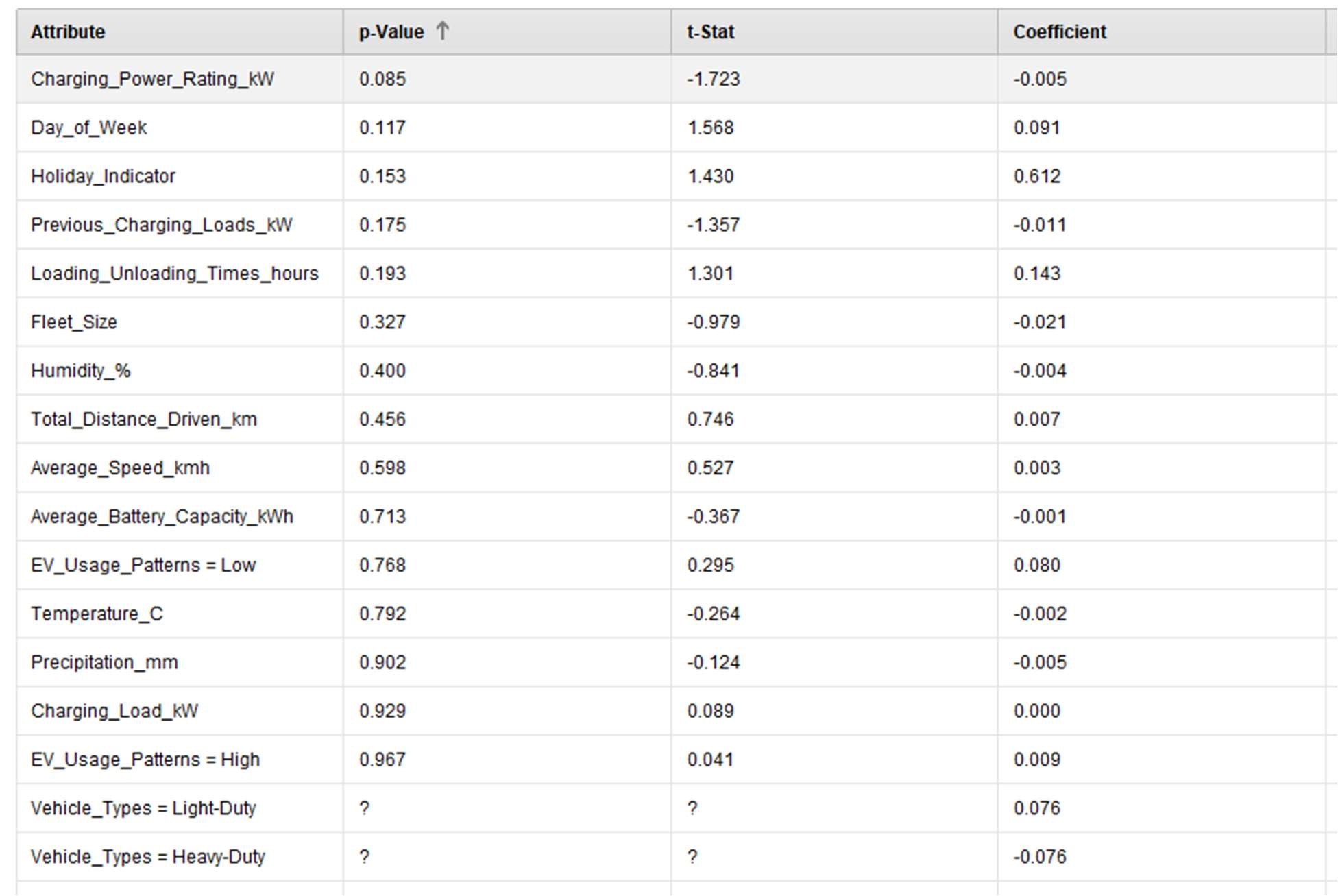
is.

When it comes to newly derived variables there are quite a few attributes that can help to quantify the data recorded. At the datasets first glance, the EV\_Usage\_Patterns appeared to the data point that would be most important when attempting to determine grid usage for a particular area. After further analysis the timestamps that were just viewed as durations but was then realized that these values can be used for time-series forecasting to help with the assessment of current preferred charging times based on current habits.

For variable selection we preferred to keep all included variables that came with the data set as it was determined that each attribute had a reasonable possibility to be a contributing factor (regardless of effectiveness) and though our model each level of contribution could be determined. As we are looking to view current stresses/loads on existing infrastructure, being able to view current grid demand can be used to help determine future expected loads as popularity and ease of access to EVs improve as technology continues to improve.

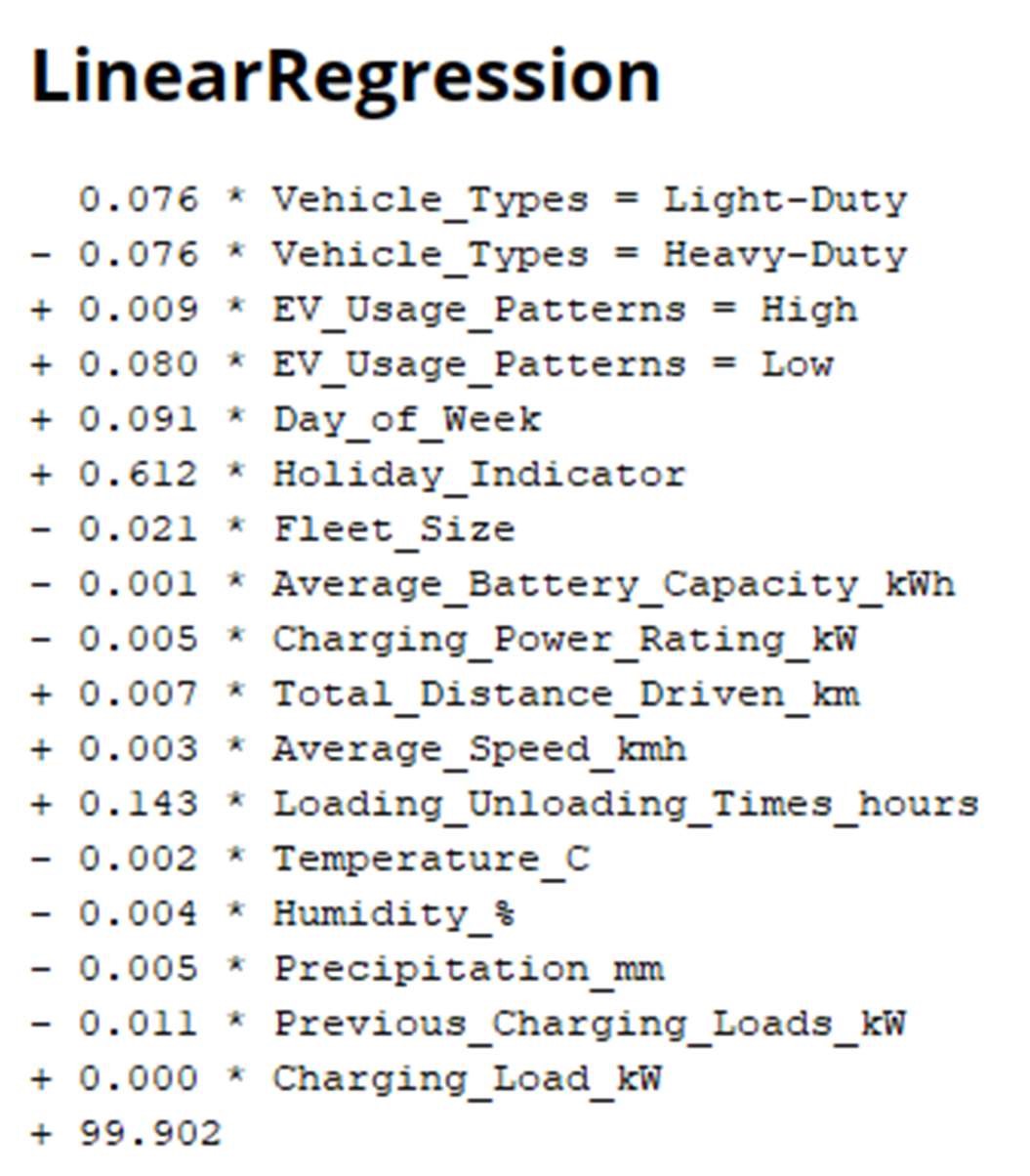
4. Discuss and justify model assessment and selection.

For this dataset we had chosen to perform a regression analysis in attempt to predict the Grid\_Demand\_MW (in megawatts). This would be imperative for understanding the relationship between EV charging and grid load balancing. Looking at the model’s performance compared to the training data there was less than reputable results when looking at the predicted values. While LR can be used for classification, with it being able to predict continuous values compared to categorical our predicted outcomes were not a precise as we were hopefully to return.



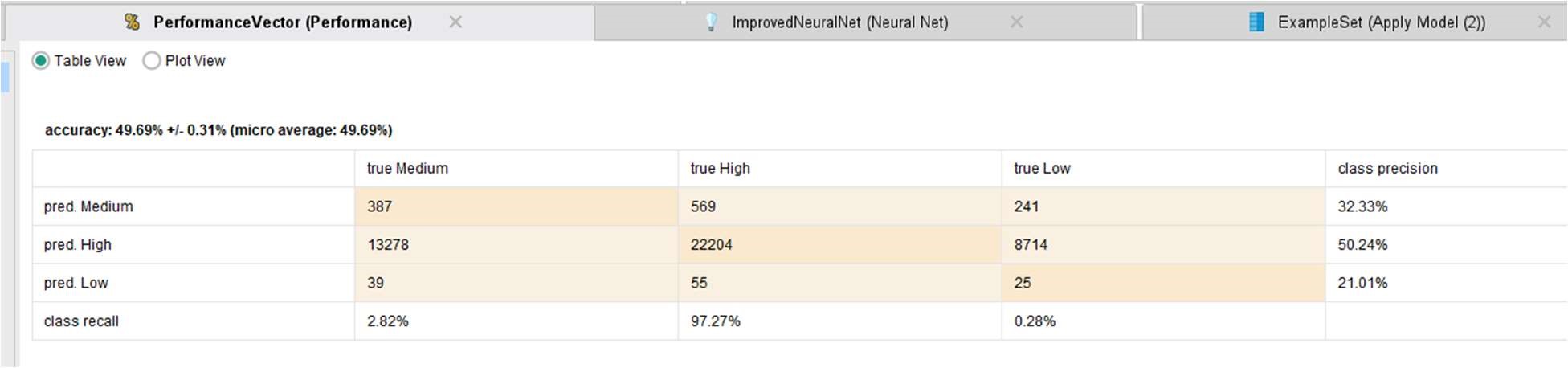
Looking at the chart above, we can visualize which attributes appear to share a stronger correlation between the recorded values with fleets activity levels showing the highest relation to grid demand with charging power rating showing the weakest relationship between attributes.

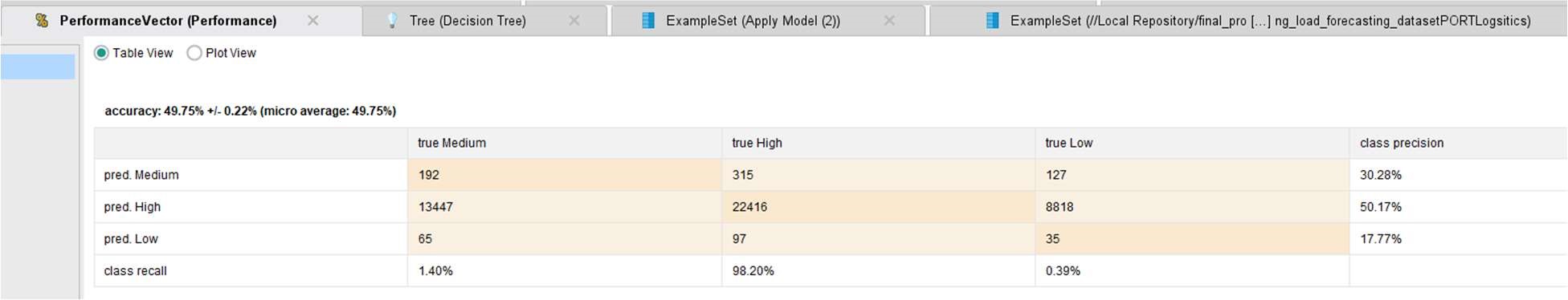
The analysis above resulted in the following equation that can be used to find the resulting grid demand in megawatts:



Alternatively, we can use (2) other models to better predict which attributes would have a higher impact on the grid as it currently exists. This would be through a Neural Network, or Decision Tree.

When comparing both models’ performance, there are many similarities between the two. My suggestion for which model to choose would learn towards the DT model as this result shows the most promise when looking to predict if a particular case will positively or negatively effect the current grid demand. Unfortunately, the figured for both (while similar) do not place relatively high accuracy or precise results prompting users with the question are EVs currently a large strain on the current grid? With their popularity increasing exponentially over the past 10 years it would appear that at least within the greater Dallas Texas area that the areas resources appear to be growing at a similar rate to that of drivers transitioning to the use of alternative fuel sourced vehicles.





Electric Vehicle Charging Patterns

This dataset provides an analysis of electric vehicle (EV) charging patterns and user behavior. It contains over 1,300 samples of charging session data, including metrics such as energy consumption, charging duration, and vehicle details. Each entry captures various aspects of EV usage, allowing for various use cases when aiming to create predictive models. https://www.kaggle.com/datasets/valakhorasani/electric-vehicle-charging-patterns

At first glance looking at the dataset, its ability to provide useful information one way or another for a main underlying question was not immediately apparent. But though further investigation, this datasets usability was beginning to show promise. With attributes such as distance driven, user type and charging duration are key components when it comes to understanding electric vehicle charging behaviors with the intent to develop predictive models related to energy consumption and user patterns.

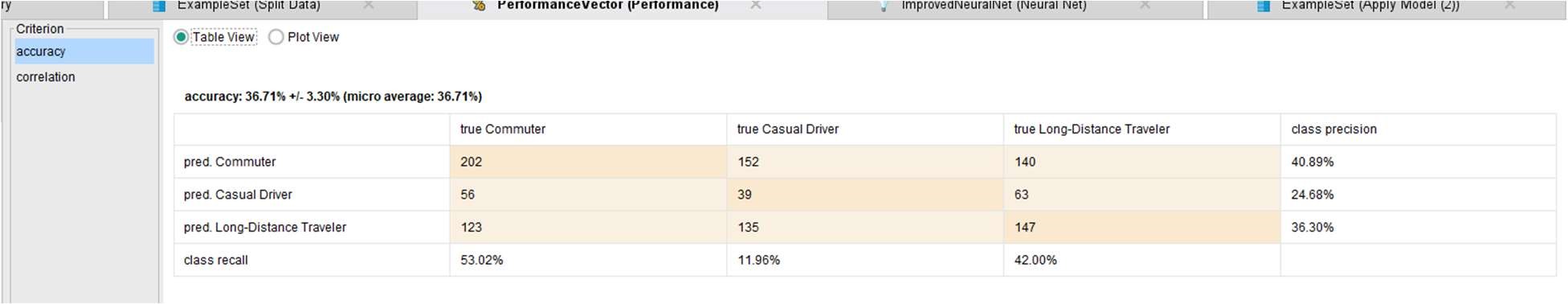
This dataset was not too raw when going thought its attributes. Some values such as charging station ID and Location seemed redundant for our use case but if users wanted to group IDs with locations, they could then create a heat map to see which location has more infrastructure in place to provide resources for their constituents.

For this data set there were many floating variables that would need to be modified to be used for DT or NN analysis. For the topic of missing variables there were some missing entries that needed to be filled, as our data set is limited in terms of entries, we did not want to omit these recordings and chose to have that substitute the average values for that attribute.

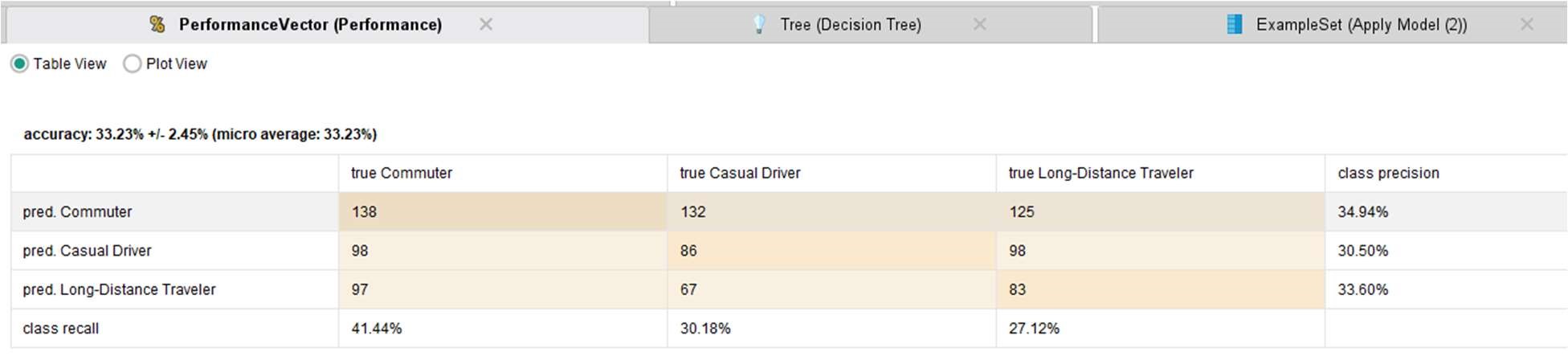
Our hope with this dataset is to be able to create a model that can accurately determine what type of drivers are using these stations (are they casual, commuter or long-distance drivers). With this knowledge we can then see what the preferred times (time of day, day of the week) for them are to charge, on a typical stop how long might they charge (are they staying until full recharge or just until they have enough to make it to their next destination). Based on this information we are looking to see if based on the recorded information if there is any correlation between these recorded sessions and the Charging Cost for the session.

For this dataset we chose to perform a Classification (prediction) model to be able to more accurately determine the weights of the supporting features and which are likely to carry more influence. A Neural Network was chosen as it can be highly flexible and capable of modeling intricate, non-linear relationships between input features and output classes. Whereas with Decision Trees that work by recursively splitting data based on feature values. DTs are relatively easy to understand, require little data preprocessing, and can handle both numerical and categorical data. Our aim is to see if we can accurately measure values recorded from various users in order to determine the type of driver that they might be. What are the features that play a pivotal role in determining this?

Neural Network



Decision tree



Looking at the results for both classifications both are not inspiring by any stretch of the imagination. From my perspective I would contribute this to a lack of or insufficient data.

Relatively speaking to the other data sets we had chosen for this exercise there were just north of 1300 recorded values this when compared to a dataset containing more that 60,000 could greatly affect our results. Looking at both options, I would suggest using the NN model as it statistically is able to correctly predict the correct outcome more often than any other attribute.

Data Cleaning and Preparation

a) Station Data Cleaning:

- Dropped rows with missing values in the `distance` field.

b) Data Merging:

- Attempted to merge charging station data with utility data based on zip codes; no matches were found.

c) Derived Variables:

* Created average utility rate features by aggregating commercial, industrial, and residential rates. However, these were not usable in models due to merging issues.

Variable Selection

The variables selected for modeling were:

* Features: `chargeTimeHrs`, `average\_rate\_non\_iou`, `average\_rate\_iou`
* Target: `kwhTotal` (total kWh consumed per session)

These variables were chosen to evaluate the influence of charging duration and utility rates on total consumption. However, due to data merging issues, further refinement was necessary.

Model Assessment and Selection Regression Models Attempted

a). Decision Tree Regression:

* Used to capture non-linear relationships between features.
* Model parameters: Max depth = 5.

b) Linear Regression:

* Simpler approach for capturing linear relationships between features and target.

Model Performance

* Due to the lack of valid data samples after merging utility data, model training faced limitations and could not produce meaningful results.
* Decision Tree and Linear Regression models were tested, but performance metrics were not available due to data insufficiency.

1. Conclusions from Model Assessment

The research underscores the complex relationship between the increasing adoption of electric vehicles (EVs) and its implications for electricity distribution and grid demand. The integration of charging behaviors, utility rate structures, and grid infrastructure presents unique challenges. Key insights from the analysis reveal many Charging Patterns and Grid Load including but not limited too frequent but low kWh charging sessions dominate, emphasizing the need for infrastructure to support short-term, high-turnover charging stations.

Utility rates may influence charging behaviors, while introducing challenges with data integration (e.g., unmatched zip codes) limit the ability to draw actionable insights. This begins to tie into some predictive modeling challenges. Regression and classification models, including decision trees and neural networks, struggled with accuracy due to data insufficiencies and integration issues.

While this study identifies promising trends and areas for improvement, the lack of comprehensive, high-quality data hinders the ability to fully assess and forecast EV impacts on the grid. The minimal integration between charging data and utility rate data highlight imperfections data availability challenges. This creates many future analyses require improved data matching, additional data sources, or alternative approaches to assess the impact of utility rates.

1. Recommendations for Management
2. Improved Data Integration and Quality

To address the challenges associated with data discrepancies, a concerted effort is required to enhance data collection, integration, and sharing among stakeholders. Utilities, charging station operators, and policymakers must standardize data formats to facilitate interoperability.

Geographic mismatches, such as zip code inconsistencies, can be mitigated through the use of Geographic Information Systems (GIS), which allow for spatial alignment of datasets. Enhanced data integration would provide a more accurate understanding of how utility rates and charging behaviors impact grid demand, enabling more effective predictive modeling and resource allocation.

1. Strategic Infrastructure Investments

Infrastructure development should focus on addressing "charging deserts" in underserved areas, particularly low-income and rural communities. Ensuring equitable access to EV charging stations will support widespread adoption and prevent disparities. Additionally, investments in high-speed charging technologies are critical to reducing the time vehicles spend at charging stations, thereby increasing the turnover rate and minimizing congestion. Urban areas with high EV adoption rates should prioritize fast chargers, while rural areas may benefit from a mix of fast and slow chargers to accommodate diverse needs. Strategic placement of infrastructure will also aid in balancing grid loads by distributing demand geographically.

1. Behavioral Incentives

Utilities and policymakers should design incentive programs to influence EV owner behavior in ways that support grid stability. Dynamic pricing models, where electricity rates are lower during off-peak hours, can encourage EV owners to charge their vehicles at times of reduced demand. Such incentives can alleviate peak-hour grid strain while promoting more sustainable energy consumption patterns. Furthermore, offering government subsidies or utility-backed rebates for home charging equipment can offset initial costs for EV owners, making electric mobility more accessible, especially to low-income households.

1. Advanced Modeling Techniques

Enhancing the accuracy of predictive models requires incorporating additional data sources beyond charging sessions and utility rates. Integrating data on weather patterns, traffic flows, and regional economic activity can provide a more holistic view of factors affecting EV usage and grid demand. Employing ensemble modeling techniques, which combine the strengths of neural networks and decision trees, can capture both non-linear and categorical relationships in the data. This hybrid approach would allow for more robust predictions, guiding infrastructure planning and policy decisions more effectively.

1. Policy and Collaboration

Policymakers must work closely with utility companies and private sector stakeholders to create a cohesive framework that aligns EV growth with grid capacity and sustainability goals. Regulations mandating off-peak charging incentives or requiring EV-ready infrastructure in new residential and commercial developments can play a pivotal role in managing grid demand.

Public-private partnerships should also be explored to share the financial burden of expanding

EV infrastructure, leveraging private investment to complement public funding. Collaborative efforts among government agencies, utilities, and charging providers can ensure that the transition to electric mobility is equitable, efficient, and environmentally sustainable.

By addressing these areas, stakeholders can effectively manage the growing EV ecosystem, balancing grid stability with sustainable energy practices and equity in access to resources.

Works Cited

Electric Vehicle Charging Patterns. (n.d.). Retrieved from Kaggle: https://www.kaggle.com/datasets/valakhorasani/electric-vehicle-charging-patterns

EV Intelligent Port Logistics. (n.d.). Retrieved from Kaggle:

https://www.kaggle.com/datasets/datasetengineer/ev-intelligent-port-logistics