

# **Classification and Trends: Distribution of Electric Vehicle Types in Washington State**

Name: Semiu Kolapo

Student Number: 501145293

Supervisor: Tamer Abdou

Date: April 1st 2024

# Why Electric Vehicle?

- Government of Canada committed to achieve 100% zero-emission vehicle sales by 2035
- Washington State's progressive policies
- Tesla

Using ML to identify trends and distribution



TESLA

# Predicting Electric Vehicles



- Can we predict type of Electric Vehicle?
- Can we predict the distribution of BEVs and PHEVs?
- Which ML algorithms can yield the best and most accurate results?
- Are the parameters within the Dataset that can give insights on the distribution of BEVs and PHEVs?
- How confident are we in our findings?

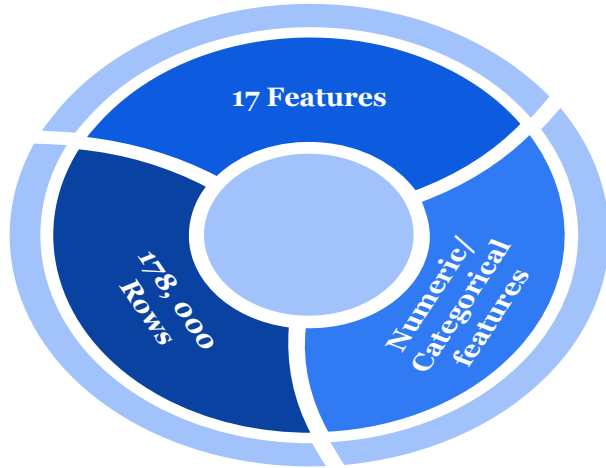
# Initial Dataset

- Electric Vehicles in Ontario – By Forward Sortation Area
  - Total EVs by Forward Sortation Area (FSA)
- Insufficient Information
- High Bias
- Feature Engineering Constraints
- Lack of Diversity

# Selected Dataset

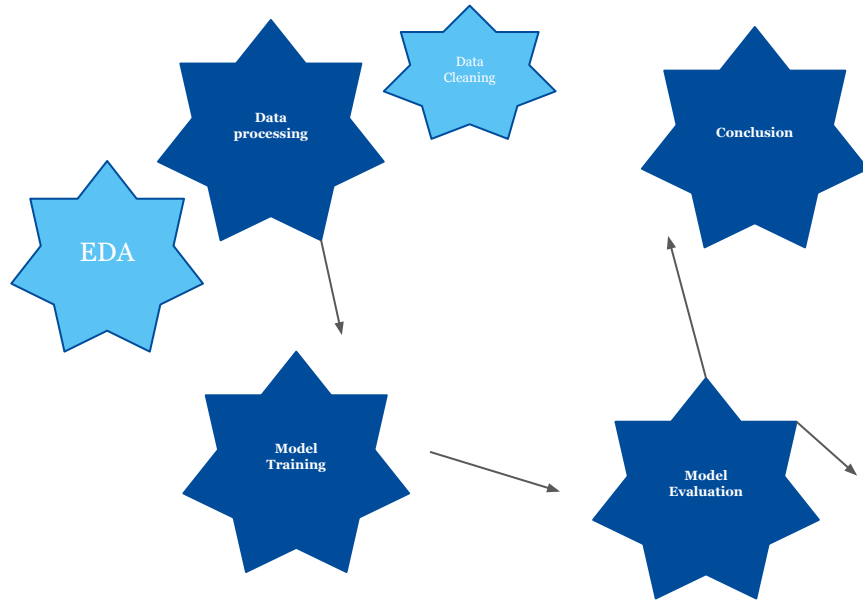
[link](#)

## Electric Vehicle Population Data

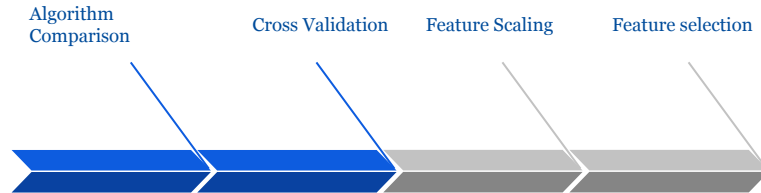


VIN	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type
Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location	Electric Utility	2020 Census Tract	CAFV	

# Approach



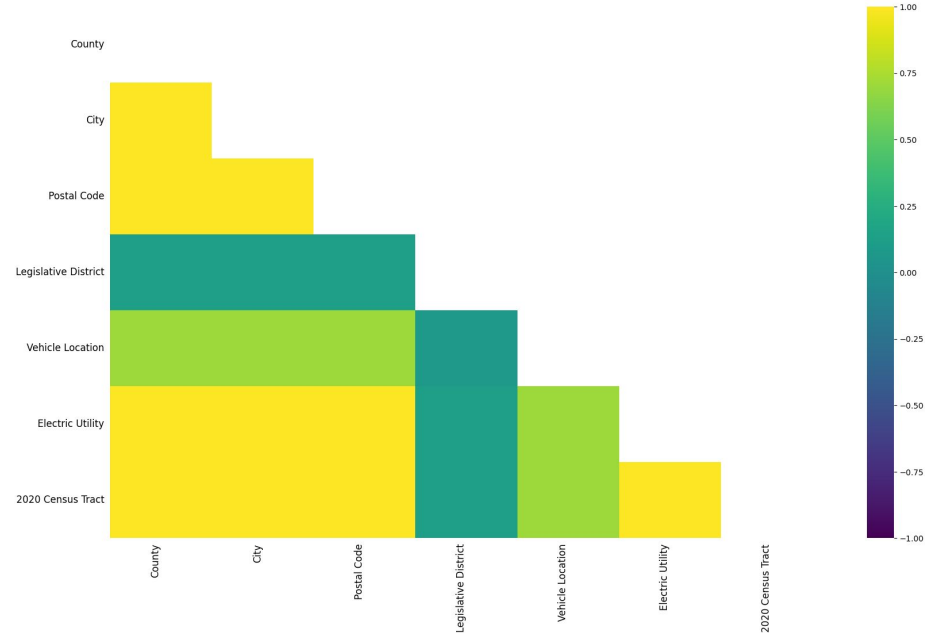
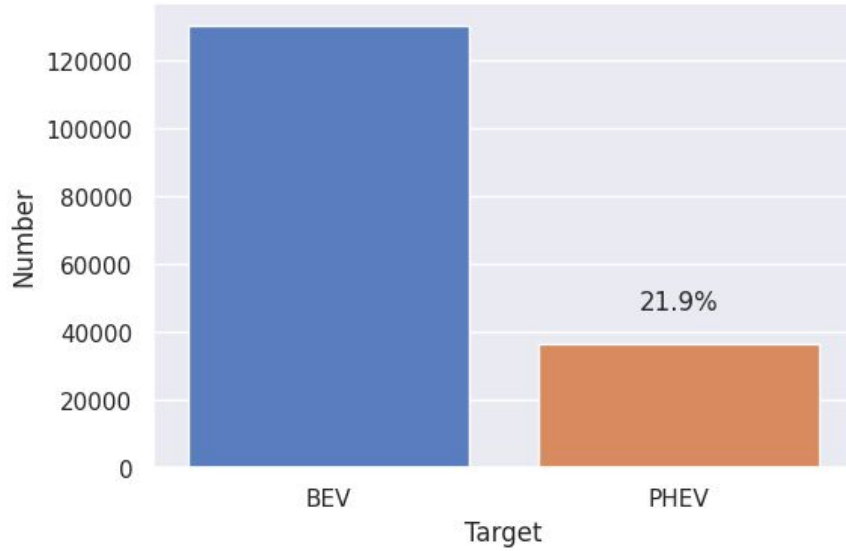
**Naive Bayes**  
**Logistic Regression**  
**Random Forest**  
**XGBoost**



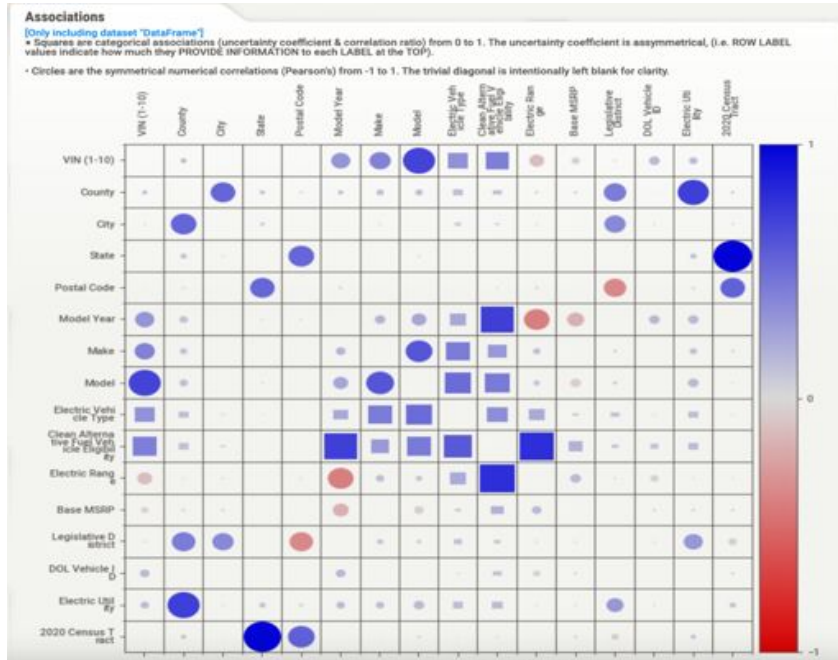
# EDA

## Missing Values

Number of Electric Vehicles Based on Target



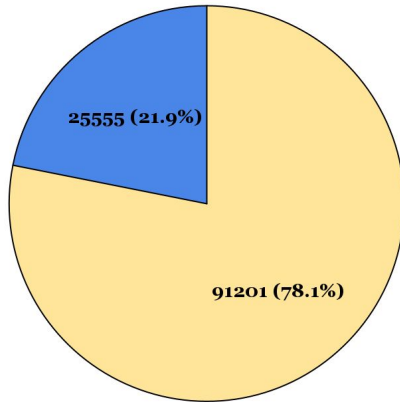
# EDA using Sweetviz and Pandas Profiling Report



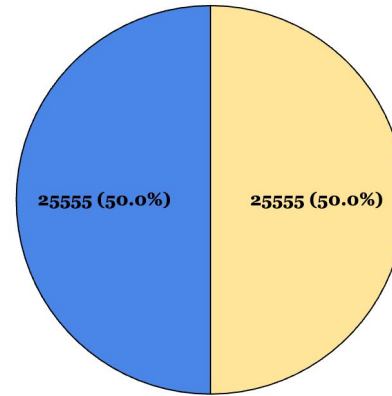


# Class Imbalance

Electric Vehicle Type



Electric Vehicle Type



NearMiss approach is utilized due to its capacity to focus on selecting relevant majority class samples near the decision boundary, effectively reducing class imbalance while preserving vital information.

# Initial Result

	Model	Accuracy	ROC-AUC
<b>0</b>	Random Forest	0.998801	0.999994
<b>1</b>	XGBoost	0.998941	0.999983
<b>2</b>	Naive Bayes	0.780491	0.727559
<b>3</b>	Logistic Regression	0.516097	0.514732

# Feature Importance

	Feature	Importance
<b>0</b>	Electric Range	0.453681
<b>1</b>	Clean Alternative Fuel Vehicle Eligibility	0.196643
<b>2</b>	Model	0.160677
<b>3</b>	Make	0.052106
<b>4</b>	VIN (1-10)	0.040262

	Feature	Importance
<b>0</b>	State	3.712071e-08
<b>1</b>	Electric Utility	9.238881e-05
<b>2</b>	Base MSRP	3.686047e-04
<b>3</b>	Legislative District	7.104148e-04
<b>4</b>	Latitude	8.808175e-04

# Final Result

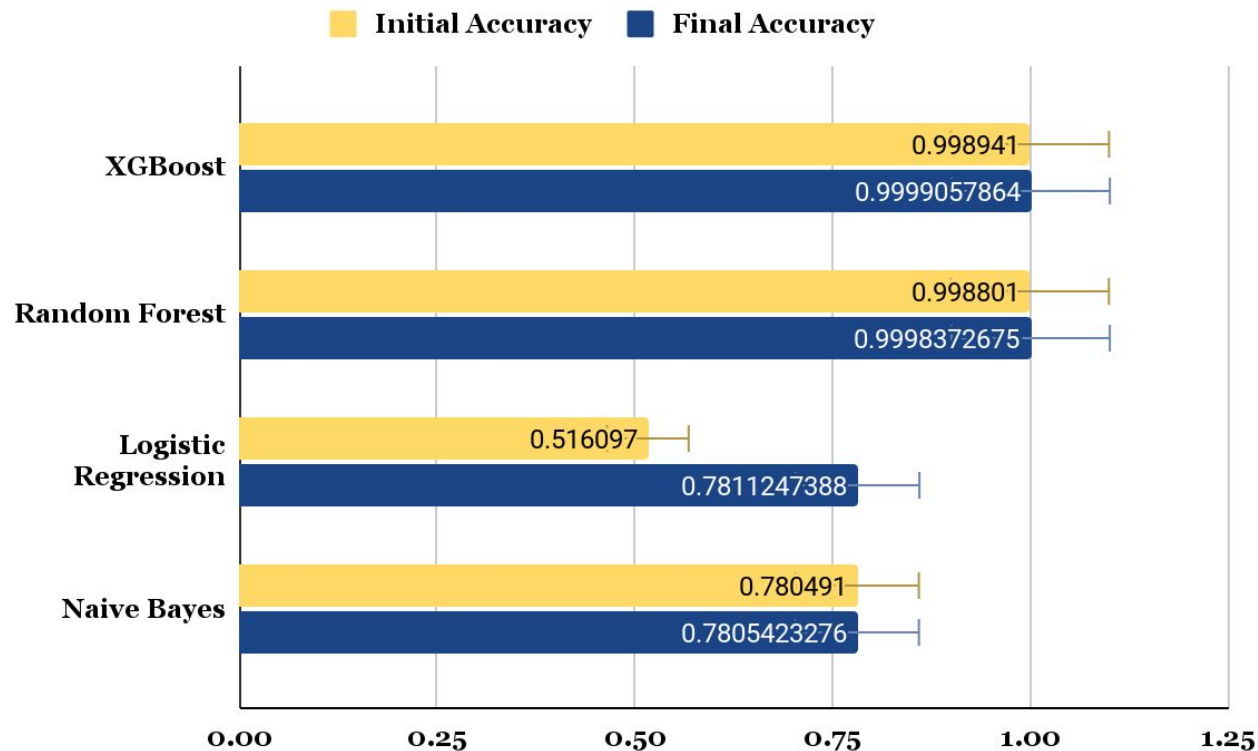
## Cross Validation

5-fold cross-validation to evaluate the performance of all the machine learning algorithms.

**Using 5-fold cross-validation strikes a balance between accuracy and computational efficiency**, ensuring that the model has enough data for training and evaluation while also generalizing well to different subsets of the dataset.

	Model	Accuracy	Accuracy w/ C-valid	ROC-AU C	ROC-AU C w/C-valid
<b>0</b>	XGBoost	0.991386718 3596794	0.99990578 64263935	0.99432234 9043133	0.99986927 74101849
<b>1</b>	Random Forest	0.98207398 22938108	0.99983726 74637706	0.988427011 3907274	0.99986766 79820891
<b>2</b>	Logistic Regression	0.781130718 0399288	0.781124738 771455	0.5	0.5
<b>3</b>	Naive Bayes	0.780411279 2022223	0.78054232 75891603	0.50065678 95225318	0.50069752 38197879

# Model Evaluation



## Interpretation

- After feature selection (drop), feature scaling XGBoost performed with the greatest accuracy of 0.99139
- After Cross Validation XGBoost accuracy increased to 0.9999059

# Conclusion

- Can we predict the type of Electric Vehicle?
- BEVs vs PHEVs
- Which Machine Learning yield the most efficient and accurate result?
- Does Cross Validation within ML yield improve results?
- What can we do to improve our findings?
- Electric Vehicle distribution by country?
- Most common make?
- Using Machine Learning it is possible to predict the type of Electric Vehicle and its distribution
- XGBoost and Random Forest are the best performed among the ML algorithms that were selected.
- C-Validation increased all the ML algorithms
- Hypertuning, feature selection, handling categorical features
- King County stands out as the epicenter of electric vehicle adoption
- Tesla - its impact in future production

# Questions?