Prediction of Alternative Fuel Vehicle Adoption

2017 NHTS Case

Sami SHOKER Ben ILBOUDO Senou AHOUNOU

Aix Marseille School Of Economics

July 29, 2022

1 Introduction

2 Data

3 Methodology

4 Results

5 Conclusion

1 Introduction

2 Data

3 Methodology

4 Results

5 Conclusion

Research Objectives

- Explore the Influencing Factor on AFV adoption.

Research Objectives

- Explore the Influencing Factor on AFV adoption.
- Predict the individual's behavior to adopt an AFV (Binary Classification).

Research Objectives

- Explore the Influencing Factor on AFV adoption.
- Predict the individual's behavior to adopt an AFV (Binary Classification).
- Compare between parametric and non parametric ML models predictive performance.

1. Introduction
2. Data
3. Methodology
4. Results
5. Conclusion

Variables

- Data Source : National Household Travel Survey.

Variables

- Data Source : National Household Travel Survey.
- Number of Observations : 20160.

Variables

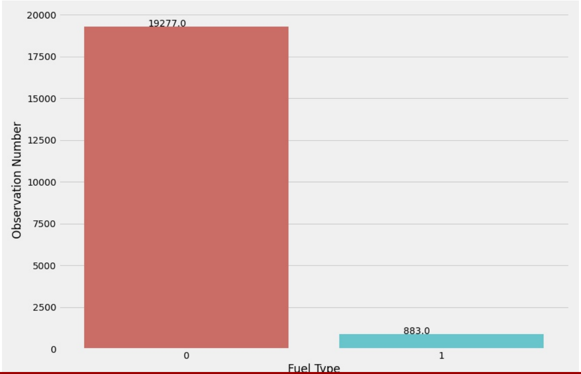
- Data Source : National Household Travel Survey.
- Number of Observations : 20160.
- Target Variable : Fuel Type (AFV).

Variables

- Data Source : National Household Travel Survey.
- Number of Observations : 20160.
- Target Variable : Fuel Type (AFV).
- Features : Household and Individual Characteristics (10 numerical 22 categorical).

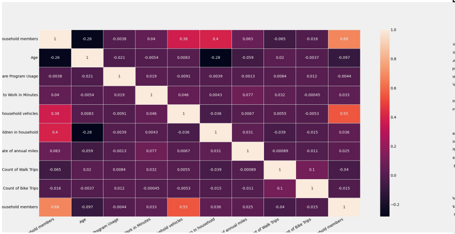
Variables

- Data Source : National Household Travel Survey.
- Number of Observations : 20160.
- Target Variable : Fuel Type (AFV).
- Features : Household and Individual Characteristics (10 numerical 22 categorical).

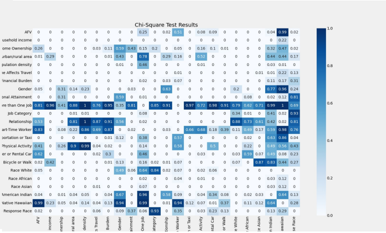


Correlation Analysis

Pearson Correlation



Chi-Square Test



Variance Inflation Factor

	Feature	VIF
0	Household income	19.60
15	Count of household vehicles	8.01
2	Count of household members	13.78
22	Count of adult household members	22.23
4	population density	10.41
...	...	...

1 Introduction

2 Data

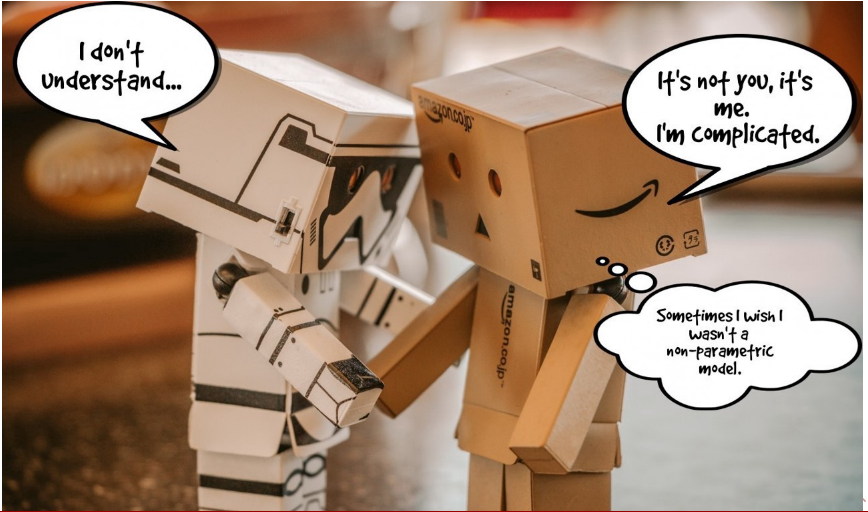
3 Methodology

4 Results

5 Conclusion

Machine Learning Models

Comparison Between Parametric non-Parametric ML models



Machine Learning Models

Parametric Models

- Logistic Regression.
- LR with FAMD.
- Ridge Regression.
- Lasso Regression.
- Logistic GAM (Semi-Parametric).

Non-Parametric Models

- Naive Bayes.
- K Nearest Neighbors.
- Neural Network.
- Support Vector Machine.
- Random Forest.

Process

- Predict the models using the original DATA.
- Create Artificial Observations (SMOTE NC) on training set only.
- Predict the models after SMOTE NC.
- Add polynomial features to linear models to test their impact on performance.
- Compare between model's performance .

1. Introduction
2. Data
3. Methodology
4. Results
5. Conclusion

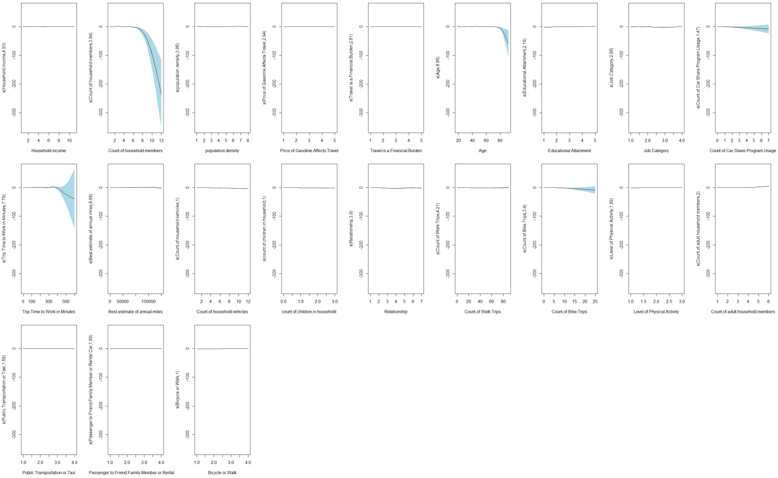
Before SMOTE NC

Classifiers	Accuracy	Recal	F1	Precision	AUC
Logistic regression	0.96	0.00	0.00	0.00	0.50
Lasso regression	0.96	0.00	0.00	0.00	0.50
Ridge regression	0.96	0.00	0.00	0.00	0.50
Naive Bayes	0.95	0.00	0.00	0.00	0.50
KNN	0.92	0.07	0.06	0.06	0.51
Neural Network	0.05	1.00	0.08	0.04	0.51
SVM	0.96	0.00	0.00	0.00	0.50
Random Forest	0.96	0.00	0.00	0.00	0.50

After SMOTE NC

Classifiers	Accuracy	Recal	F1	Precision	AUC
Logistic	0.73	0.47	0.13	0.07	0.60
Polynomial Logistic	0.96	0.00	0.00	0.00	0.50
Logistic with FAMD	0.53	0.67	0.11	0.06	0.60
Lasso	0.72	0.45	0.12	0.07	0.59
Polynomial Lasso	0.78	0.36	0.12	0.07	0.58
Ridge	0.72	0.45	0.12	0.07	0.59
Polynomial Ridge	0.79	0.37	0.13	0.08	0.59
Logistic GAM	0.67	0.54	0.12	0.07	0.61
Naive Bayes	0.52	0.68	0.11	0.06	0.60
KNN	0.74	0.29	0.09	0.05	0.52
Neural Network	0.06	0.97	0.08	0.04	0.49
SVM	0.29	0.71	0.08	0.04	0.49
Random Forest	0.90	0.12	0.09	0.07	0.53

Logistic GAM Plots



1 Introduction

2 Data

3 Methodology

4 Results

5 Conclusion

Conclusion

- All models suggest no discrimination before SMOTE NC (AUC 0.5).

Conclusion

- All models suggest no discrimination before SMOTE NC (AUC 0.5).
- Models Performance Increased after SMOTE NC.

Conclusion

- All models suggest no discrimination before SMOTE NC (AUC 0.5).
- Models Performance Increased after SMOTE NC.
- Linear models outperformed non-parametric models.

Conclusion

- All models suggest no discrimination before SMOTE NC (AUC 0.5).
- Models Performance Increased after SMOTE NC.
- Linear models outperformed non-parametric models.
- Polynomial features didn't help increasing model's performance.

Conclusion

- All models suggest no discrimination before SMOTE NC (AUC 0.5).
- Models Performance Increased after SMOTE NC.
- Linear models outperformed non-parametric models.
- Polynomial features didn't help increasing model's performance.
- No strong non-linear relationship between target variable and features.

Conclusion

- Overall model's performance is low (even after SMOTE NC).

Conclusion

- Overall model's performance is low (even after SMOTE NC).
- Household individual's characteristics aren't sufficient to predict the consumer's behavior to adopt an AFV.

Conclusion

- Overall model's performance is low (even after SMOTE NC).
- Household individual's characteristics aren't sufficient to predict the consumer's behavior to adopt an AFV.
- Features about AFV industry can also impact the individual's behavior such as battery range, price etc (Literature Review).

Conclusion

- Overall model's performance is low (even after SMOTE NC).
- Household individual's characteristics aren't sufficient to predict the consumer's behavior to adopt an AFV.
- Features about AFV industry can also impact the individual's behavior such as battery range, price etc (Literature Review).
- Suggestion for future work to take into account features from all different aspects mentioned above.

Thank You