# EVALUATING ELECTRIC VEHICLE USER MOBILITY DATA USING NEURAL NETWORK BASED LANGUAGE MODELS

Kevin Alvarez (North Carolina State University), Emerson Wenzel (Tufts University), Arielle Dror (Smith College), Omar Isaac Asensio* (Georgia Tech)

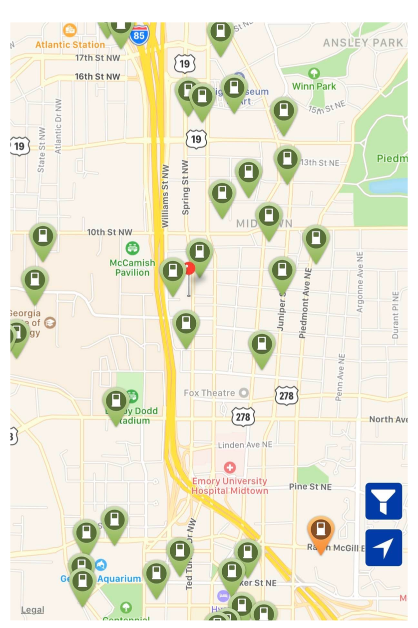
Georgia Institute of Technology, Civic Data Science REU

## Background

Bloomberg estimates that by 2040, 55% of vehicle sales will be electric. To support the increasing number of drivers, we need a strong infrastructure to support them. There is research that quantifies current electric vehicle (EV) infrastructure, but there is no research that investigates the quality of the infrastructure. This leads us to our research question:

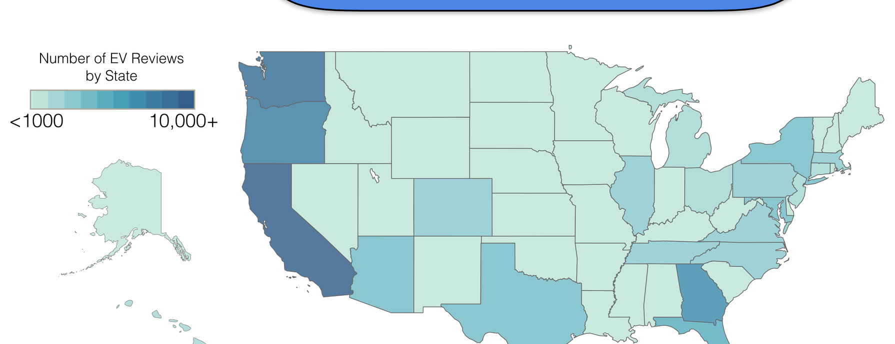
***How well is the current EV infrastructure serving drivers?***

## Data



*From 2011-2015…*

- 140,183 Reviews
- 29,532 Users
- 17,731 Locations in U.S.




We obtain our data from the leading app for crowdsourced electric vehicle charging station data. Of the 140,183 reviews in the dataset, 127,257 are from the United States.

## Additional Data Acquisition

For POI Classification:

Google Places

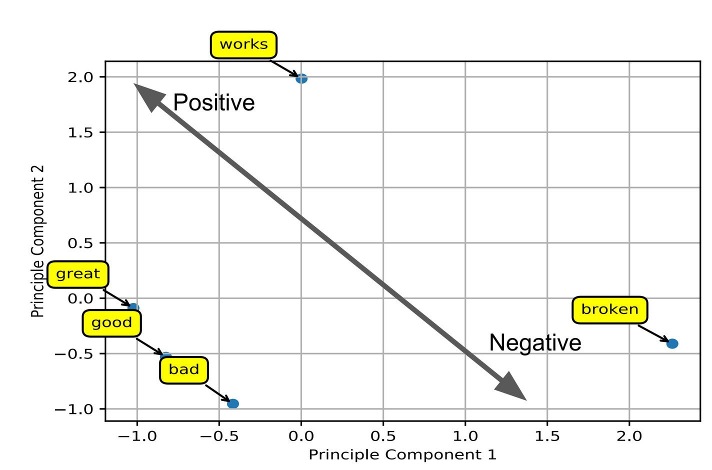
In order to prepare for our analysis, we collected more data to allow us to form different station groupings of interest.

For Urban/Rural Classification:

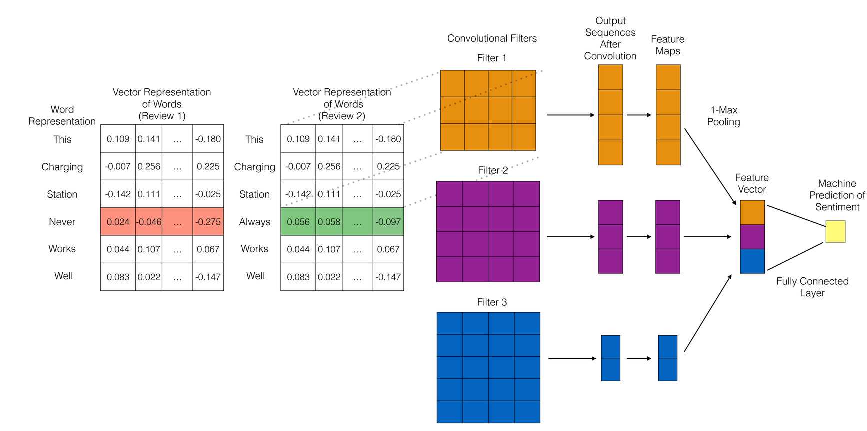
United States Census Bureau

## Machine Learning

It would take a prohibitive amount of time to hand-classify these reviews. Machine learning allows us to analyze this dynamically growing data set.



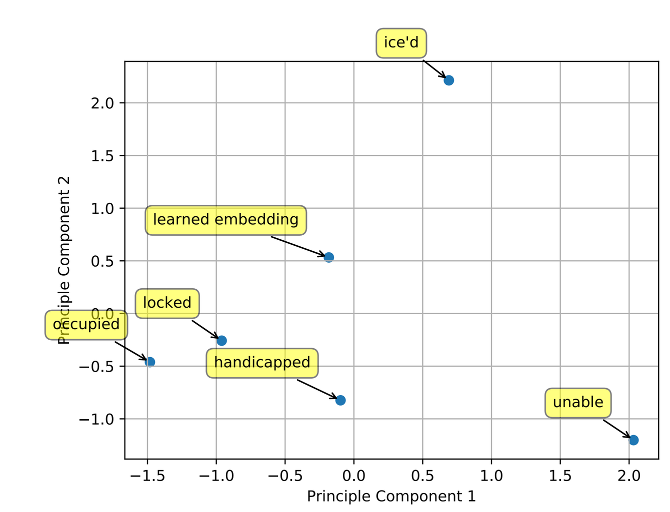

We use pre-trained word2vec word embeddings to represent review words in a distributed manner that captures semantics. Similar words are nearby each other in this vector space, and linguistic concepts are encoded in the dimensions of the space.



MODEL PERFORMANCE

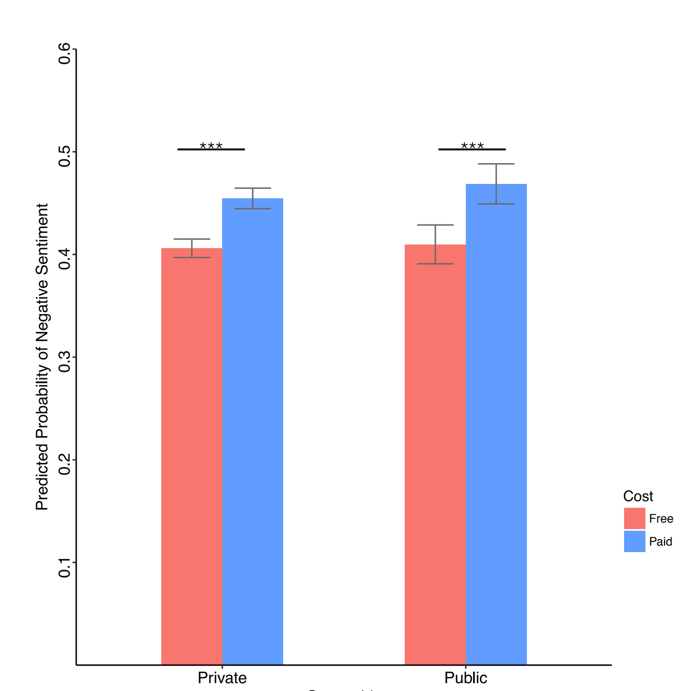
| Model | Accuracy (Percent) | Precision | Recall |
|---|---|---|---|
| **CNN** | **84.1** | **0.85** | **0.86** |
| SVM | 76.5 | 0.78 | 0.79 |
| LR | 78.5 | 0.79 | 0.82 |

**CNN:** Convolutional Neural Network
**SVM:** Support Vector Machine
**LR:** Logistic Regression



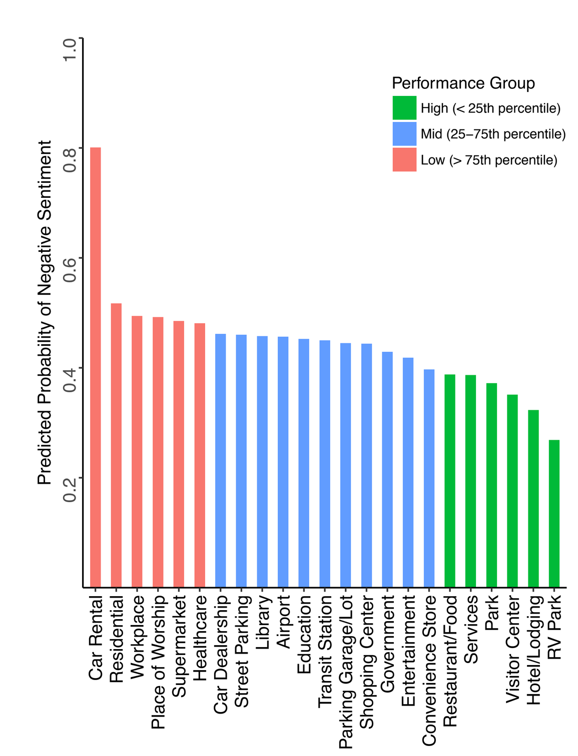

The CNN is able to learn domain specific terminology.

## Results



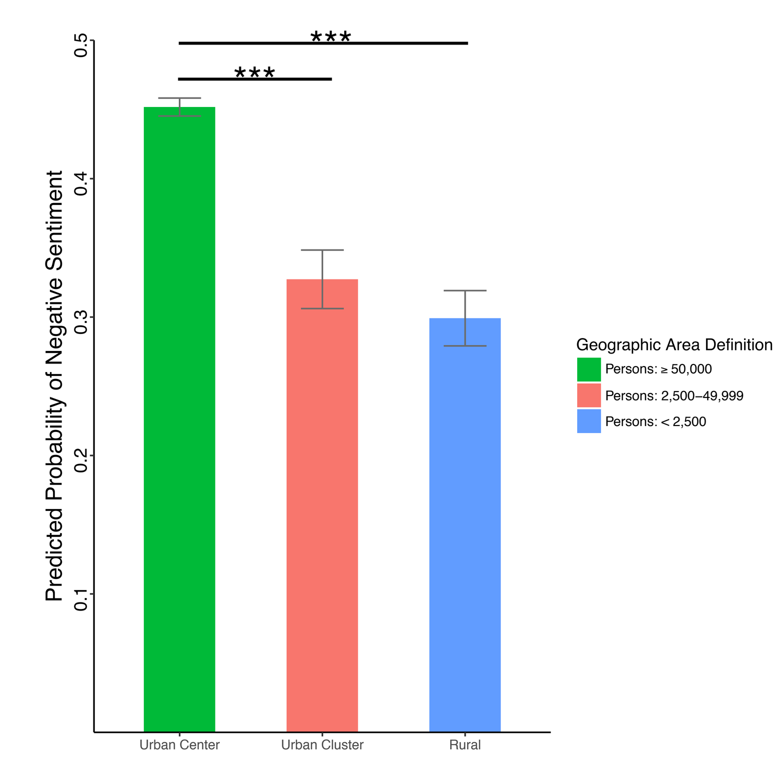

**Negative Sentiment by Ownership and Cost**

Although we do not find any statistically significant differences between publicly and privately owned stations, we find that free stations outperform paid stations in both groups.




**Negative Sentiment by POI**

We find that there are statistically significant differences between all combinations of high, mid, and low performing stations.




**Negative Sentiment by Urban/Rural**

We find that stations in urban areas underperform compared to other areas. We also find that there are no statistically significant differences between performances in urban cluster and rural areas.

## Statistical Analysis

**Measuring Negative Sentiment**

To measure negative sentiment, we calculate the probability of negative sentiment at the *location level*:

$$P_{(negative|location_i)} = 1 - \frac{count(positive\ reviews_i)}{total\ reviews_i}$$

**Two-Part Fractional Response Model (FRM)**

First stage models factors driving selection to review. Second stage models the effects of station characteristics on the negativity score.

| | Review Rate | | | Negativity Score | | |
|---|---|---|---|---|---|---|
| | *FRM* (I) | *FRM* (II) | *FRM* (III) | *FRM* (IV) | *FRM* (V) | *FRM* (VI) |
| **Geographical Area** | | | | | | |
| Urban Center | -0.021** | -0.038*** | -0.039*** | 0.149*** | 0.123*** | 0.122*** |
| | (0.008) | (0.007) | (0.007) | (0.013) | (0.012) | (0.012) |
| Non-Urban | 0.025** | 0.016* | 0.016 | 0.016 | 0.004 | 0.004 |
| | (0.012) | (0.010) | (0.010) | (0.016) | (0.014) | (0.014) |
| **Type of Location** | | | | | | |
| Public | -0.010 | -0.012 | | 0.010 | 0.008 | |
| | (0.013) | (0.015) | | (0.015) | (0.012) | |
| **Station Characteristics** | | | | | | |
| Number of Connectors | -0.082*** | -0.074*** | -0.074*** | -0.011*** | -0.005 | -0.005 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Number of Networks | -0.011 | -0.012 | -0.012 | 0.030* | 0.020 | 0.020 |
| | (0.014) | (0.017) | (0.017) | (0.016) | (0.016) | (0.016) |
| PlugScore | | -0.042*** | -0.042*** | | -0.058*** | -0.058*** |
| | | (0.002) | (0.002) | | (0.002) | (0.002) |
| POI Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustered SE | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Observations | 127,257 | 127,257 | 127,257 | 127,257 | 127,257 | 127,257 |
| $R^2$ | 0.117 | 0.235 | 0.235 | 0.049 | 0.120 | 0.120 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

## Discussion

We have a number of theories as to why these differences exist:

***Ownership/Cost:***

Although we do not find any differences between publicly and privately owned stations, we do find a difference between free and paid stations in both groups. This could be because users who pay to charge have greater expectations from performance.

***POI:***

**Active/Passive Management:** Stations where managers take an active role in their upkeep have better performance.
**Travel:** Stations at POI related to travel perform better because people are more desperate for a charge.
**Incentives:** POIs can bring traffic to their business by installing and up-keeping a charging station.
**Normative Behavior:** Stations where users follow established norms have higher performance.

***Urban/Rural***

Stations in urban areas might underperform because of congestion.

## Acknowledgements

We would like to thank:

- Dr. Christopher Le Dantec and Dr. Ellen Zegura, Co-Directors of the Civic Data Science REU for their guidance throughout the program.
- Dr. Amanda Meng and Dr. Carl DiSalvo for their feedback.
- Francella Tonge for her support.

This project was supported by grants from the Anthony and Jeanne Pritzker Family Foundation, Sustainable LA Grand Challenge and the National Science Foundation.

* Principal Investigator