INL/EXP-23-73777-Revision-0

SULI Oral Presentation

August 2023

Jason Patrick Clifford

Changing the World's Energy Future

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

INL/EXP-23-73777-Revision-0

SULI Oral Presentation

Jason Patrick Clifford

August 2023

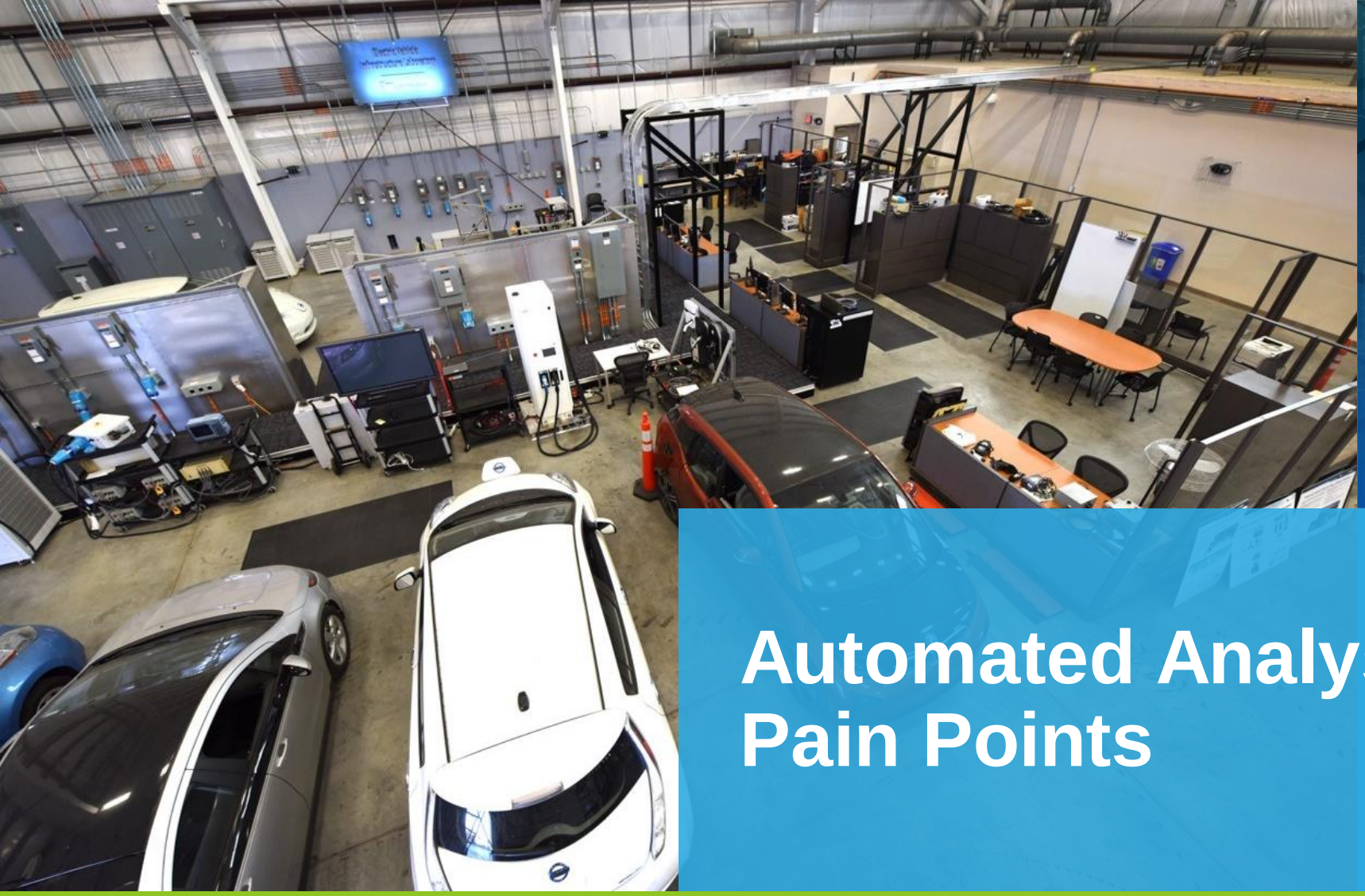
Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Abstract

Understanding the prevalence and severity of issues that customers face when charging their electric vehicles (EVs) is crucial in order to improve the charging experience across the United States. This project utilizes web-scraping, machine learning (ML), and natural language processing (NLP) techniques to analyze and categorize user-generated reviews. Selenium was used to build a data collection tool that can scrape vast amounts of user review data from the PlugShare website. Sentiment analysis was employed on this dataset in order to filter out negative reviews for further analysis. NLP techniques such as tokenization and word embedding were then used to convert user-written comments into a numerical format that an ML model can interpret. Multiple ML approaches are currently being explored in order to identify and categorize the charging issues being talked about in each review. Ultimately, the results from the ML model will be visualized and explained in a report on customer pain points to be delivered to the ChargeX Consortium, therefore revealing specific areas for improvement in the customer charging experience.



August 2nd, 2023

Jason Clifford

SULI Intern, Electric Vehicle Infrastructure

Automated Analysis of Customer Pain Points

Idaho National Laboratory

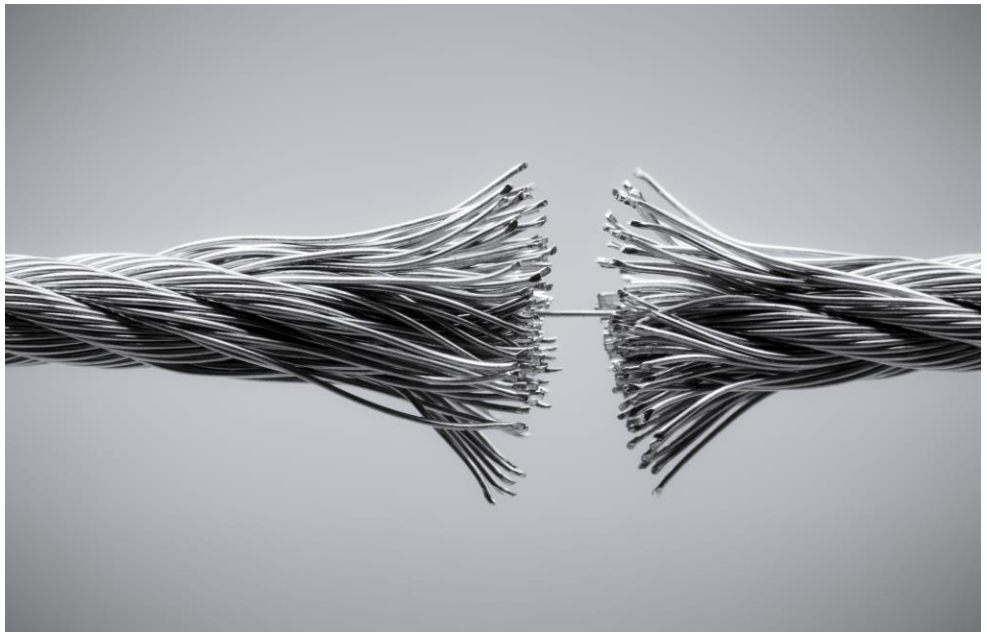
Overview

- Introduction
- Approach and Methodology
- Data Collection Process
- Utilizing Machine Learning
- Expected Deliverables
- Next Steps

Introduction

- Reliable charging infrastructure is essential to sustaining the growth of electric vehicle (EV) use.
- This project aims to gain insight into the challenges faced by EV owners across the United States.
- We are analyzing reviews from PlugShare, a popular EV route planning website/app, to understand the prevalence of common charging issues.

Common Customer Pain Points



1. Had to try multiple chargers
2. No charge delivered
3. Lower power output than expected
4. Payment Issue
5. Took multiple attempts to connect
6. App sign-up required
7. Charge stopped mid-session
8. Errors with mobile app
9. Charger restarted after contacting service
10. Restricted access to charger
11. Connector/cable issues
12. Touchscreen issues
13. Misleading charger availability status
14. Long wait time

Approach and Methodology

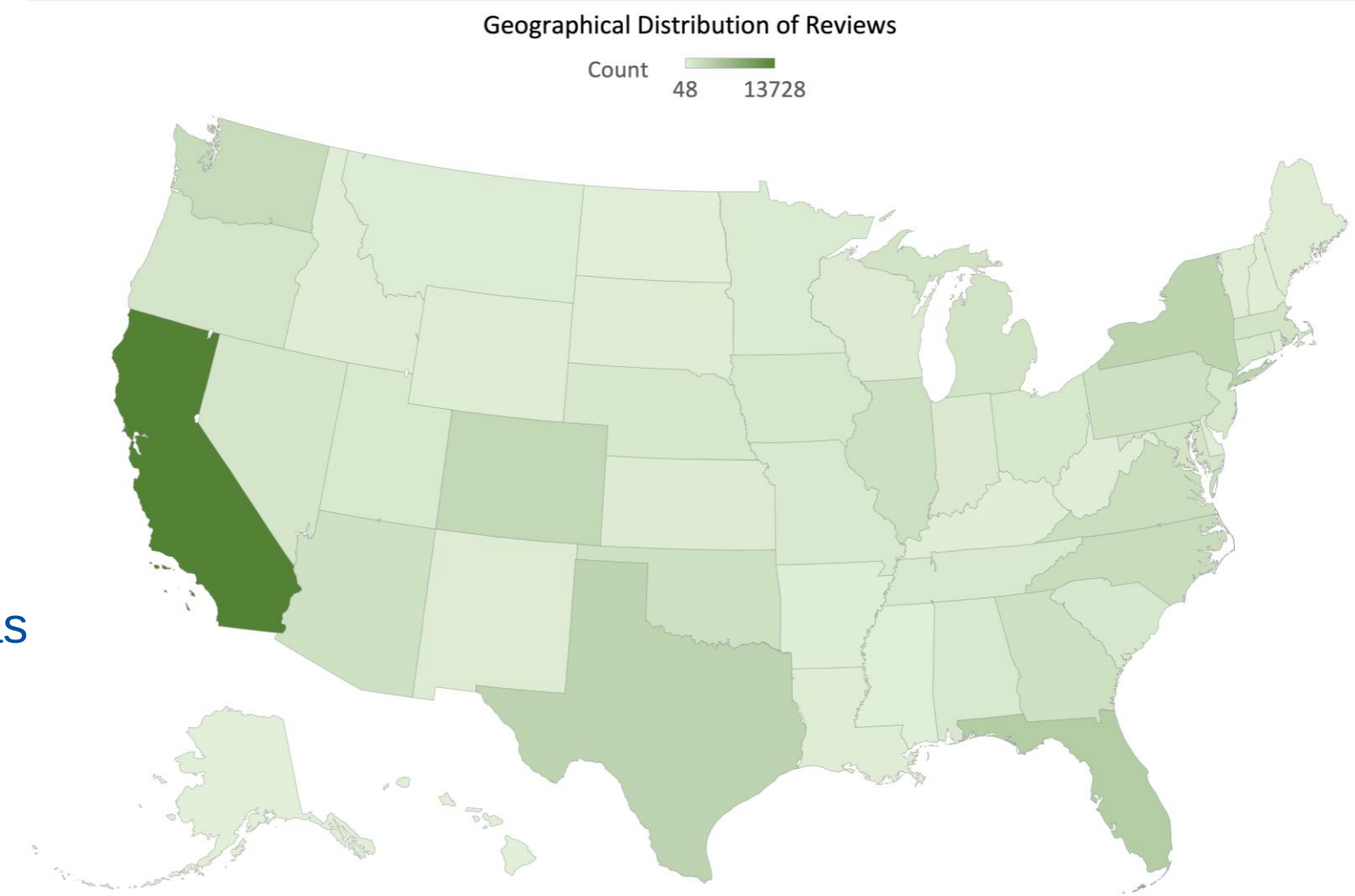
- An automated web-scraping tool pulls the recent review data from every high-power charging station in the country.
- A machine learning model is being built which aims to analyze this dataset and utilize natural language processing to identify what sort of issues the customer is talking about in each review.
- This analysis should allow us to observe the prevalence of various charging issues in relation to region, station operator, time of year, and vehicle model.

Web-Scraping Tool

- PlugShare IDs were manually collected from each station's page on the website.
- Due to the structure of the website, the scraper can only access data from the 50 most recent reviews at each station.
- Individual lists of IDs for each state were created, so that current data can be collected on a state-by-state basis.

Overview of Dataset

- **72,314 reviews** from **3,161 stations**
- Data was collected from every 70+ kW station in all 50 states as well as Puerto Rico and D.C.



Utilizing Machine Learning

- Three machine-learning (ML) approaches are being explored for this project, each with their own pros and cons

1. Supervised Learning
2. Unsupervised K-means Clustering
3. Fuzzy C-means Clustering

- Natural Language Processing (NLP) techniques are also being used to transform the user-generated comments into a structured format that can be analyzed by the ML model.

Supervised Learning

- Manually label the pain points present in each comment in a training set of reviews

Pros:

- Able to define the problem categories ourselves
- Will yield easily interpretable results

Cons:

- Process of labeling training set is time-consuming
- May overfit the training set due to the highly variable nature of review comments

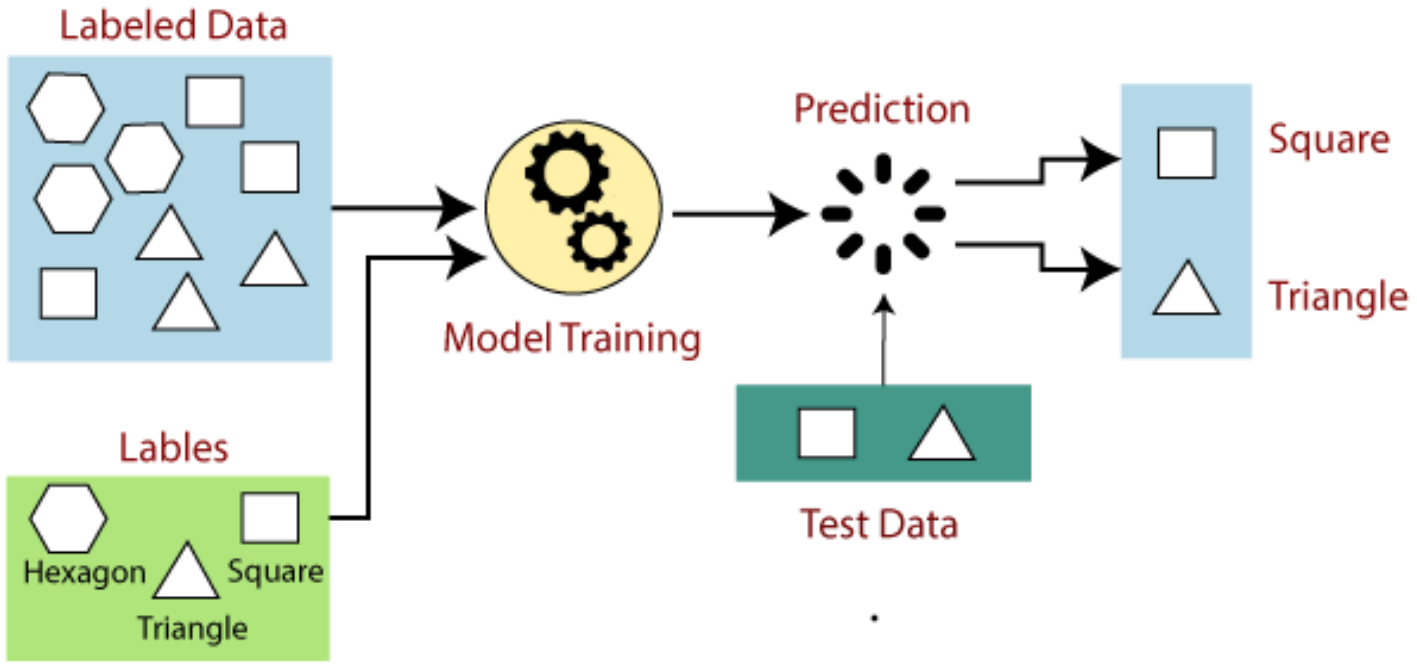


Image from "Supervised Machine Learning" by Javatpoint. Available at https://www.javatpoint.com/supervised-machine-learning (accessed July 18th, 2023)

Unsupervised K-means Clustering

- Model sorts each review into one of various clusters based on its relation to the contents of other reviews

Pros:

- Does not require labeled data
- Could discovered unexpected problem categories

Cons:

- Need to specify the number of clusters
- Only allows each review to be a part of a single problem category
- May struggle to create meaningful clusters due to noisy data

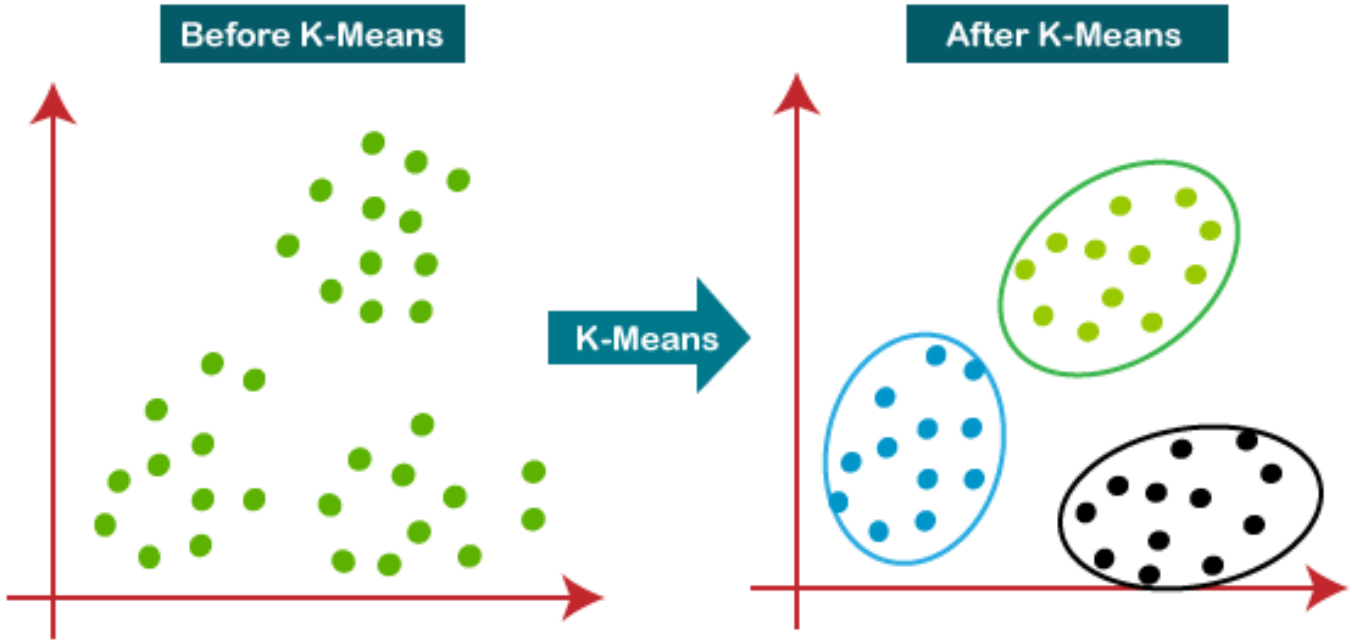


Image from "K-Means Clustering Algorithm" by Javatpoint. Available at https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning (accessed July 18th, 2023)

Fuzzy C-means Clustering

- Instead of assigning each review to a cluster, it provides a membership value to every possible cluster for each review

Pros:

- Does not require labeled data
- Would account for reviews which contain more than one type of charging issue

Cons:

- Most computationally expensive approach
- Most challenging to interpret results
- Will face similar struggles to K-means

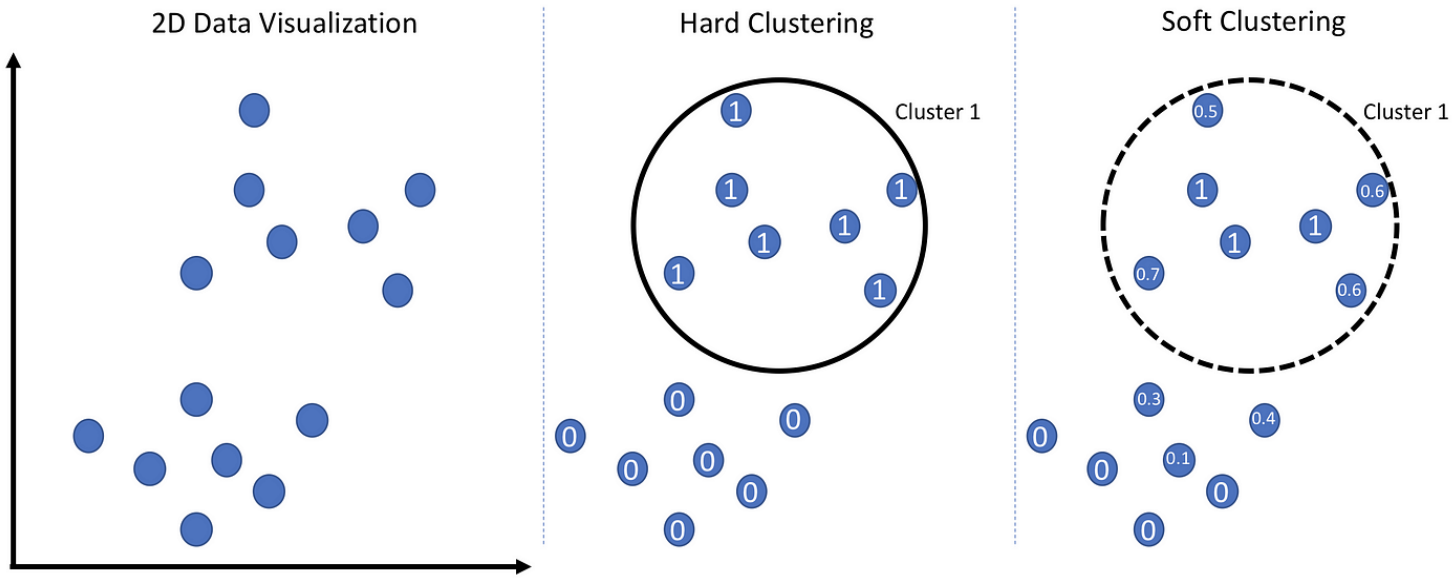
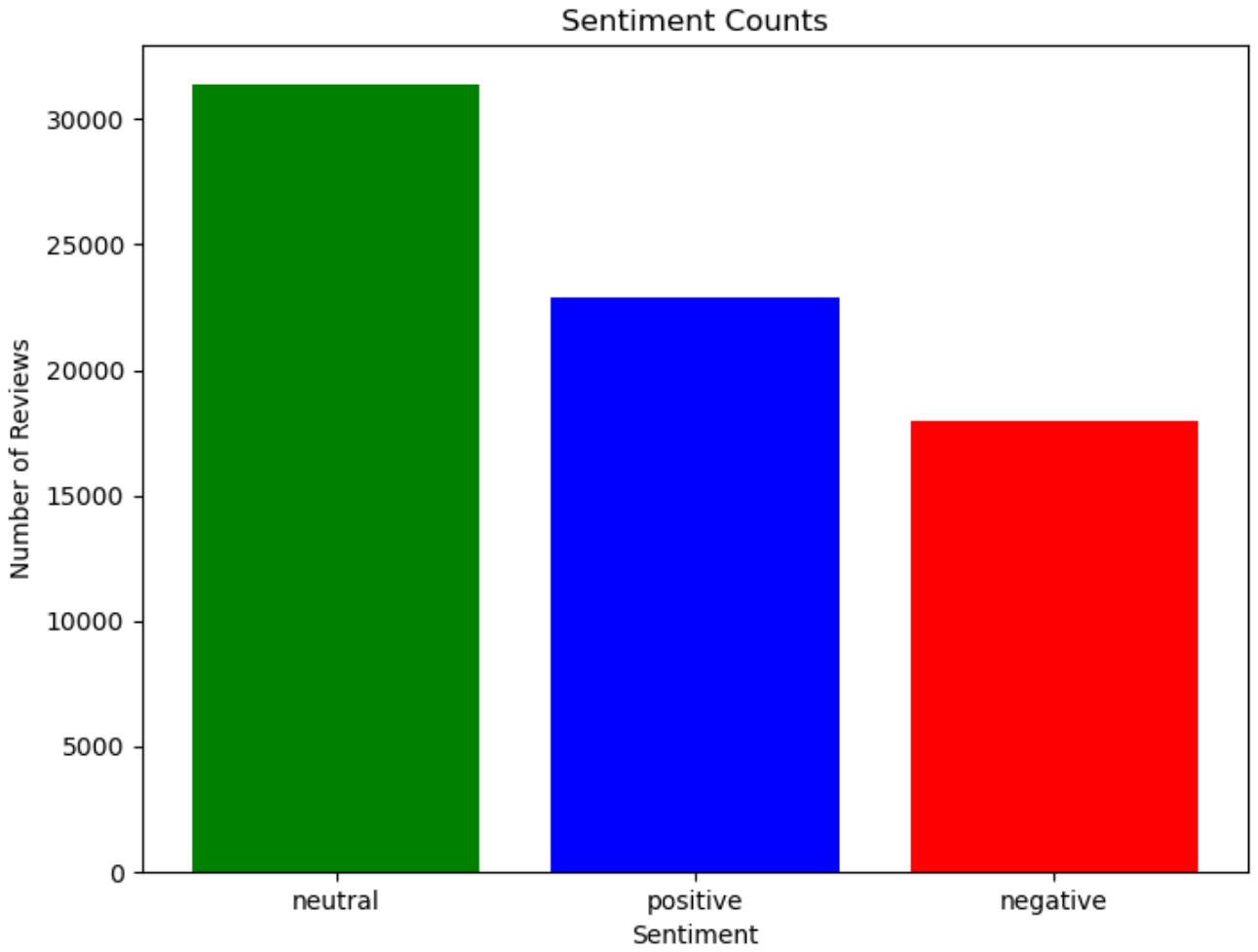


Image from "Fuzzy C-Means Clustering with Python" by Yufeng. Available at https://towardsdatascience.com/fuzzy-c-means-clustering-with-python-f4908c714081 (accessed July 18th, 2023)

Natural Language Processing

- A pre-trained **sentiment analysis** model called VADER was used to classify reviews as either positive, neutral, or negative.
- Preliminary sentiment analysis results are classifying about **25%** of scraped reviews as negative.
- The NTLK Python library was utilized to clean and **tokenize** the comments, removing unnecessary words and breaking up each comment into a list of individual words known as 'tokens'.
- The Global Vectors for Word Representation (GloVe) **word embedding** model was used to convert each 'token' into a vector representation. This essentially translates the text data into a numerical format that the ML model can understand.



Expected Deliverables

- Processed Dataset
- Data Visualization
- Code and Documentation
- Report on Customer Pain Points

Next Steps

- We are currently tuning the ML model and experimenting with the three different approaches to see what is the most effective.
- Once that is finished, visuals will be created to interpret and analyze the results of the ML model and relate the pain points present in each review to the other parameters of the dataset.
- Finally, a comprehensive report outlining our findings will be written and delivered to the National Charging Experience Consortium, or ChargeX Consortium, where it will be of great value to researchers, industry experts, consumer advocates, and others.

Acknowledgements

- **Dr. Benny Varghese,** research engineer at INL who's served as my mentor for the summer and provided his insight and guidance over the course of the internship.
- **Dr. Mayuresh Savargaonkar,** engineer at INL with a specialty in machine-learning who is assisting with the ML side of the project.
- **Idaho National Laboratory,** for providing me with this incredible opportunity!