Using machine learning techniques to aid environmental policy analysis: a teaching case in big data and electric vehicle infrastructure.

Omar Isaac Asensio[1*], Ximin Mi[2] and Sameer Dharur[3]

[1]*School of Public Policy & Institute for Data Engineering & Science (IDEaS), Georgia Institute of Technology, Atlanta, GA USA*

[2]*Georgia Tech Data Visualization Lab, Georgia Institute of Technology, Atlanta, GA USA*

[3]*School of Computer Science, Georgia Institute of Technology, Atlanta, GA USA*

[*]correspondence to: asensio@gatech.edu

Abstract

For a growing class of prediction problems, big data and machine learning analyses can greatly enhance our understanding of the effectiveness of public investments and public policy. However, the outputs of many machine learning models are often abstract and inaccessible to policy communities or the general public. In this article, we describe a hands-on teaching case that is suitable for use in a graduate or advanced undergraduate public policy, public affairs or environmental studies classroom. Students will engage on the use of increasingly popular machine learning classification algorithms and cloud-based data visualization tools to support policy and planning on the theme of electric vehicle mobility and connected infrastructure. By using these tools, students will critically evaluate and convert large and complex datasets into human understandable visualization for communication and decision-making. The tools also enable user flexibility to engage with streaming data sources in a new creative design with little technical background.

Learning Goals

1. To use data-driven tools for natural language processing (NLP) in policy relevant contexts
2. To consider ethical issues related to the performance of machine learning classifiers and cloud-based visualization to generate policy insights.
3. To leverage these automated content analysis tools in research evaluation of sustainability behavior and environmental decision-making in transportation and electric mobility.

Introduction

Given recent advances in the use of big data in government, scholars have argued for both theoretical and practical reorientations in pedagogy to meet a perceived data skills gap in the

training of public managers [1][2]. Data science, its uses and implications for society are rapidly permeating across several social science fields, with emerging examples in schools of public policy, economics, environmental studies and management [3][4][5][6]. For example, scholars in the Network of Schools of Public Policy, Affairs, and Administration (NASPAA) have previously highlighted curricular innovations in information technology-related competencies such as big data and cloud computing as a way to keep up with demands for workforce training [7], including importantly, graduate training for non-traditional students who need to understand how to use these data science tools as part of their regular employment [8]. In the context of environmental decision making, case-based instructional methods can facilitate active learning with data science tools [9]. These instructional strategies can be used to promote evidence-based policy positions and statistical analyses by actively engaging students [10] to think critically about sustainability challenges. However, these tools have historically required specialized technical knowledge, and there are as yet relatively few examples of the uses of big data in the classroom for applied policy analysis.

We document a teaching case that provides hands-on instruction on a typical big data problem in which a machine learning (ML) model is used to make predictions about performance using text as data. The challenge for students is to take the outputs of supervised text classification algorithms and then use the machine predictions to evaluate the social, policy or sustainability-related features through visualization. Active learning with visualization tools are increasingly needed to help convey evidence to practitioners and to increase comprehension among public servants [11]. This case introduces students to visualizations that help provide context for machine learning approaches by blending advances from computing into policy studies. Often times, ML algorithms generate predictions to help users understand future performance in critical social decisions [12][13]. These computational results, however, usually lack any display of insights from spatial or temporal dimensions, and therefore, fail to present to the audience stories from these meaningful dimensions. With this consideration in mind, we describe a Georgia Tech collaboration between the School of Public Policy and the Data Visualization Lab at the Georgia Tech library. We focus on the creation and use of case-based and evidence-based pedagogy in environment and sustainability [9][14]. The collaboration resulted from a need to provide students with resources to develop more engaging and appealing visualizations to tell stories hidden behind streaming data sources.

The collaborative instruction was piloted in both beginning graduate and advanced undergraduate public policy courses at Georgia Tech. The goal of the sessions was to introduce students to suitable data visualization skills, and to provide experience with cloud-based tools as a method of exploring, presenting and interpreting the results of machine learning analyses with dashboards and visual analytics. We use this hands-on approach to teach students about sustainability-related issues in the context of transportation infrastructure, where complex data sources and methods are used. This involves analysis of public charging services for electric vehicle mobility and text analysis using consumer data. Students explore automated content analysis tools and also learn to evaluate consumer sentiment and perceptions about service provision of sustainable charging infrastructure.

We first describe the background of the case in the context of electric mobility. Next, we describe the machine learning workflow, both generally and for our particular case. Following this, we describe the visualization tasks and conclude with the use of this approach to help students generate policy insights in the classroom. We also mention some practical aspects of the teaching experience.

Case Examination

As the transportation sector is now a dominant source of CO_2 emissions in the United States, displacing gasoline and diesel fuels via vehicle electrification has grown in importance. Widespread adoption of electric vehicles (EV) is expected to yield substantial public health benefits from reduced air pollution and tailpipe emissions. As a result, travel behavior and strategies to increase sustainable infrastructure have captured attention. Prior research has shown that public policies supporting electric vehicle mobility have emphasized the quantity rather than the quality of connected infrastructure [15][16][17], and it is unclear how well the existing charging infrastructure is meeting the needs of EV users. To address consumer sentiment in the public discourse on electric vehicles, a data-driven approach is presented to teach students to evaluate whether service reliability – directly related to the quality of services – could remain a critical barrier to technology adoption in this domain.

Given the large-scale use of EV infrastructure in public settings, private digital platforms such as charging station locator apps and other mobility apps are collecting real-time data on EV usage. This provides a wealth of streaming data for evaluators to process. However, consumer data such as EV user reviews from public charging stations is often unstructured and lays dormant as text. In practice, it would be costly for policy analysts or government agencies to classify this information by hand for performance assessment. As an example, at a rate of 100 reviews per hour, a human expert would take about 32 work weeks to analyze unstructured reviews at a national scale [15].

To alleviate this problem, students deploy a machine learning classifier to process EV charging station reviews automatically using social data in a digital platform and natural language processing. This approach lets students reduce processing times for impact evaluation from weeks of human annotation to just minutes of computation. Although this teaching case uses computational solutions in a specific sustainability context, the use of big data and NLP tools to evaluate a range of policy issues in environmental and natural resource policy is rapidly emerging [18][19][20][21]. The classroom practice also builds on modular aspects of data science education in social domains that includes strategies to build data acumen, incorporate real-world applications, foster interdepartmental student collaboration, consider ethics and human subjects protections, and develop curriculum suitable for distance or remote learning environments [22]. For additional relevant cases on applying machine learning and text mining strategies in contexts such as land use and mobility, see [23].

Implementing Natural Language Processing

We introduced students to the case by giving a basic overview of natural language processing without any assumed prior knowledge. The objective is to give students visibility into the steps involved in using text as data. We first described the basic elements of a text mining process for unstructured data that is broadly applicable to any machine classification task. For example, there are two known categories of machine learning algorithms – supervised and unsupervised learning. Unsupervised learning may be the most familiar to students in the social sciences. It entails forming relationships (e.g. clusters) of previously unknown patterns in data without having explicit labels. Supervised learning, on the other hand, entails learning a mapping of input data to its corresponding labels as learned from real world data. For example, short user generated texts can be labeled as class A or class B (e.g. positive or negative sentiment) for learning on an external dataset. Because many unsupervised models sometimes yield results that are uninterpretable, this case focuses on applications of text mining approaches using supervised learning.

The general pipeline for supervised learning algorithms is split into two broad phases – training and testing as shown in Figure 1. Students begin the training phase by collecting a dataset consisting of consumer EV charging station reviews that need to be classified into positive or negative sentiment. The corresponding ground-truth labels for training the learning algorithm are provided and have been previously labeled by human annotators [15]. Students randomly split the provided dataset into three subsets – the training set, the validation set and the testing set. This data split is done in order to keep a fresh subset of the data to test the learning model that has been trained during the training phase. Students learn that the typical process for deciding the split ratio between training, validation and testing is empirically determined (for example 80% training, 10% validation and 10% testing), and the data itself is randomly shuffled to avoid any possible sampling biases during the training phase. Students distinguish this process flow for classification in Figure 1, which intentionally partitions and does not use all available data. This is to be differentiated from typical regression-based analyses in causal inference problems in which the objective is to use all available data to parameterize the model. In the classroom, it was necessary to reinforce this point on differences in data process flow between ML prediction and causal inference tasks. We also found it useful to include evaluation questions in supporting discussion questions and problem sets in order to help students solidify these comparative concepts related to data processing. Students reported that debriefs on data process flows between prediction and causal inference tasks helped to facilitate learning (e.g. Assignments facilitated learning 5.0 / 5.0, Spring 2019, N=19 students).

In the classroom discussion, students also learn to identify the ethical issues and risks of having sampling imbalances in the training data, which can lead to algorithmic or observational biases, especially when analyzing subpopulations from historical data [24]. This is because many machine classification algorithms are known to propagate errors that favor the dominant labels or groups, and thus supervised learning methods often require a human in the loop. Students discuss how machine classification algorithms can often supplement but not replace human expert capabilities in many social domains (for example, in assisting judges in crime-and-sentencing decisions or resource allocations in health and safety or fire code inspections). For a recent exposition of the importance of fairness in algorithmic decision, see [25] for reference. It is

common for data scientists to train algorithms on data that is not representative of the population of interest, or is replete with historical or naturally prevalent biases. By allowing students to experiment hands-on with decisions about the training data used to build a classifier, the benefits and limitations of the tool are revealed empirically. For example, does the training data favor urban or non-urban locations, potentially leaving out review samples from certain subgroups or under-represented communities? How would this affect the appraisal of consumer sentiment for subpopulations? Students benefit from discussion and intuition about ethical and other practical implementation issues. For example, a student writes: "I enjoyed using the theory to implement practical solutions…I think these will be useful ways to collect and present data."

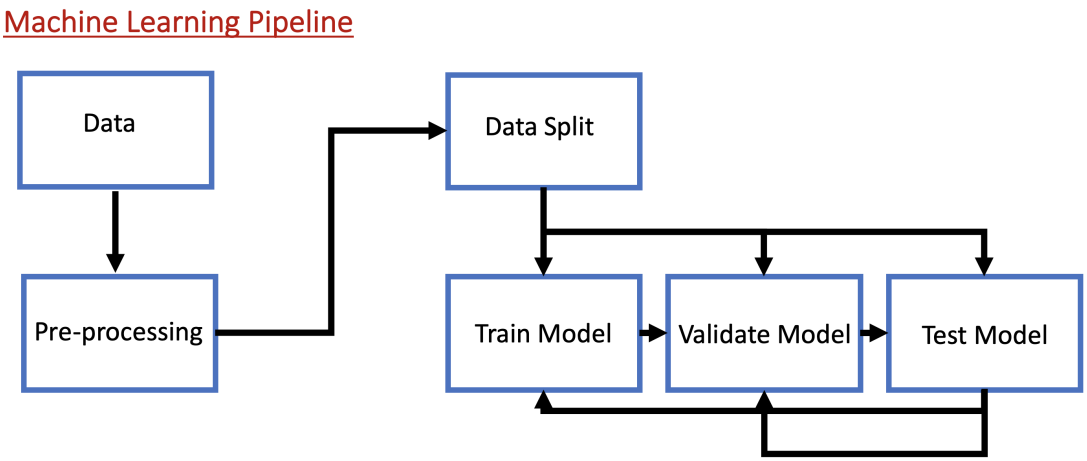


Fig 1. A general machine learning data process flow

Prior to training a model, students learn some technical details about the steps needed to pre-process data. This included a hands-on demo of coding implementations such as harmonizing data formats, removing stop words, and other tasks, which are represented in the data and pre-processing steps in Figure 1. For example, a general machine learning process flow typically involves selection and tuning of *hyper-parameters*, which are ancillary parameters needed to optimize performance. For this exercise, students build a classifier using guided protocols. They learn to classify consumer reviews using a single layer neural network based on a convolutional neural network (CNN) as the supervised learning classifier. CNNs have recently been shown to be effective in many natural language processing tasks that deal with short sentences of text and consistently outperform conventional non-neural net-based classification models such as SVM or logistic regression. For a quantitative comparison, see [15].

Students convert the input review text into pre-trained vector representations of words using the *'word2vec'* word embeddings depicted in Figure 2. The purpose of these word embeddings is to make the text data amenable to mathematical operations as word vectors of the kind required in the training process. Word embeddings closer together in the vector space are more similar to each other, whereas those that are far apart in the vector space represent concepts that are more dissimilar. These word embeddings are found to accurately represent patterns of natural language as they have been pre-trained on over 100 billion words and phrases from Google news [26]. These word vectors are then programmatically padded with null tokens to

ensure they are all of the same length, making them a matrix of representations for sentences. For convenience, this manipulation such as data cleaning, data normalization and other pre-processing steps have been completed and provided for student use using protocols described elsewhere [15]. As data science practitioners spend a significant time on data cleaning and processing, it is useful in a classroom discussion to communicate this so that students could anticipate such challenges in their future roles. More advanced students are able to modify parameters in the code and are encouraged to experiment with additional features.

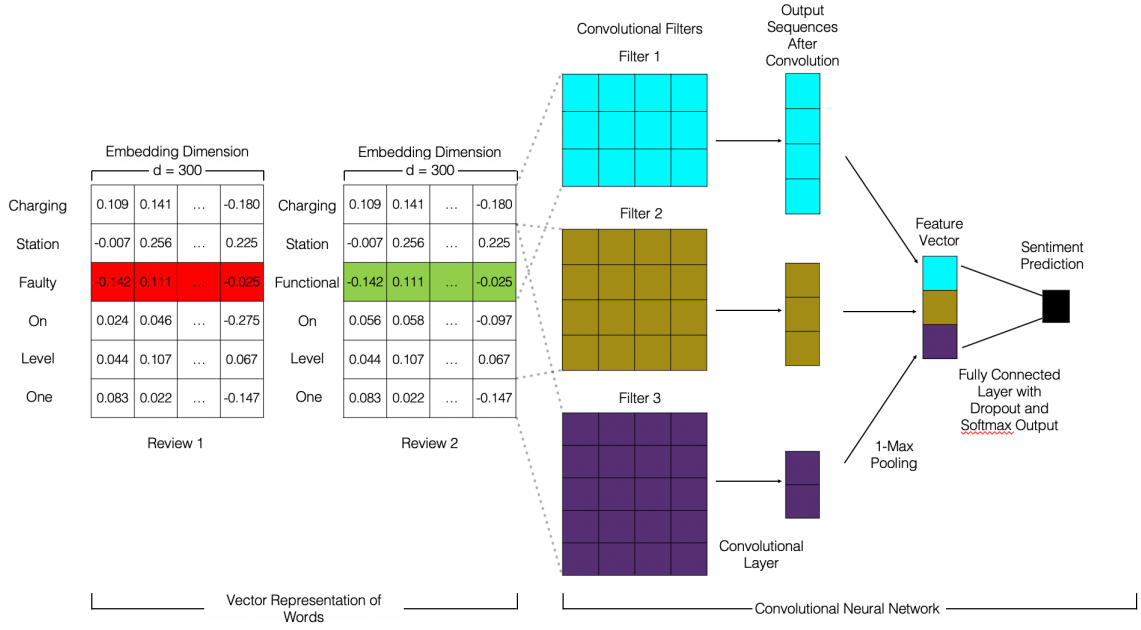


Fig 2. A typical neural network model architecture showing the vector representations of words and the CNN structure.

Another key idea for discussion in the training stage is model validation. This refers to setting aside a small percentage of the training data to be used exclusively for hyperparameter tuning. This becomes important because we want to ensure that the hyperparameters we tune our model with do not overfit to the intricacies of the specific training data and can generalize well to new data. Overfitting a model in machine classification is a very common temptation and it usually happens when the model learns the detail and noise of the training data to the extent that it performs poorly on new data.

Following the in-class demos, students were provided instructions to independently train the CNN classifier and then run the model on the test data to generate binary sentiment predictions which are used to analyze performance. By this stage, students fully execute a machine learning pipeline, which allows them to understand how various parameters impact the accuracy of the classifier predictions and decide on appropriate metrics to evaluate performance. In the interactive case, students are given sample replication code and are encouraged to experiment with various parameters. Students learn that the classifier can be iteratively improved upon based on its performance on the test data, until an optimal accuracy is reached. Constraints include factors such as such as the nature of the data, the structure of the model, and the

computational resources available. While interested students are encouraged to experiment with the hyperparameters, the case materials include default values for these hyperparameters with which optimal results are expected to be readily replicable. For additional details, students are able to reference the technical articles in [15][16][17].

In order to accommodate students from a variety of backgrounds, the sample code and implementation instructions provided in this case are carefully tailored to ensure that students can run the models off-the-shelf with little to no modification by users. In the classroom discussion, students learn that machine learning algorithms are truly no panacea to predictive problems, but can in combination with human input, extend capabilities to automatically process large volumes of text to help lower evaluation costs. Next, we describe the nature of the data and visualizations.

Mobile app data

We introduced students to a digital dataset of electric vehicle (EV) charging station reviews from a popular mobile app. This includes charger location information, unstructured user reviews (e.g. text data) including the ML generated sentiment classifications (e.g. positive and negative), and surrounding point of interest information overlaid on U.S. Census data. The dataset provides rich information that can be used to answer questions about consumer sentiment and the provisioning of EV charging services. A key objective is to secure the source data, while allowing students to access derived data from the unstructured EV charging reviews [15], with visualization capacity to introduce the underlying civic data science described previously. After the hands-on implementation of sentiment classification tasks, the next step is to make sense of the output derived data.

Streaming data and classifying EV reviews

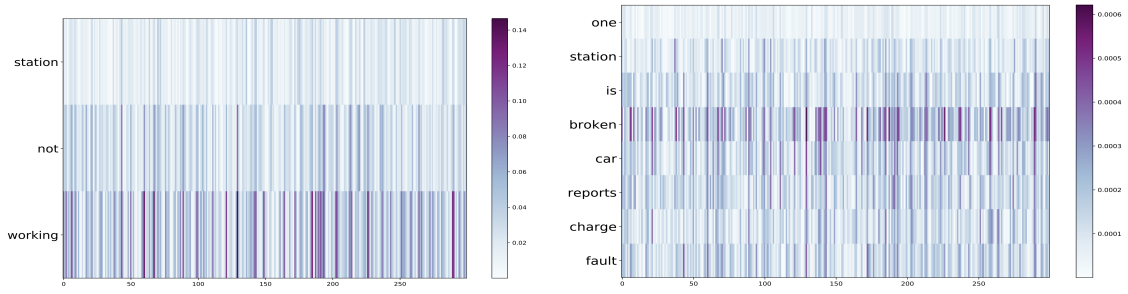
Students are given access to a dataset comprising ~25,000 electric vehicle charging stations consumer reviews in California. The objective of the exercise is twofold: (i) to build a robust and accurate classifier that captures the sentiment of the user reviews in this domain, (ii) to use this trained, generalizable classifier on a different set of given user reviews to set the stage for specific policy-relevant discussions about EV charging service provision. The use of machine classified reviews to evaluate large-scale consumer sentiment offers a lower cost computational solution for research evaluation. For the training, students process a sample of ~9000 consumer reviews that have previously been labelled by human expert annotators as positive or negative experiences. Although the full implementation details of the training data curation are outside of the scope of this case, more advanced students can nonetheless experiment with different modeling parameters and compare the technical performance for both neural-net classifiers as well as non-neural net based classifiers, which have been built as addendums in the code.

In order to familiarize all students with the tools needed to seamlessly replicate the code and promote experimentation, prior to the case, we offered a series of hands-on training sessions such as *A Beginner's Guide to Python* in partnership with the Georgia Tech Library. These data boot camps get students up and running and help them to be able to initialize and load open source packages in Python. This was a particularly important cascading option for learners to separate the basic tool setup and code navigation from the case content. We note that the

accompanying case code is intended to be user friendly and utilizes open-source tools to support the analysis (e.g. Tensorflow, Keras among others) to perform sentiment prediction using different techniques. The code also includes instructions for evaluating the performance of learning algorithms using both conventional baseline techniques for classification such as logistic regression, support vector machines as well as neural-net-based classifier based on CNNs. Although performance can be evaluated with accuracy or balance measures for the classifier, it is also helpful to visualize the results to provide a more complete picture of the ability for the predictive algorithm to discover the words that most strongly contribute to positive or negative sentiment. Having completed the demos and training sessions for machine classification, the next steps involved generating visualizations to improve model interpretability as well as spatial analysis of the consumer data.

Visualizing machine classifier performance

Given the black box criticism of many machine learning models, a key feature of our teaching case is the ability to provide students with a tool to enhance the interpretability of model's sentiment predictions. This is a pivotal area in machine learning research today, as many scholars argue that the future success of the field depends critically on algorithms being transparent enough to build user trust [27]. We instructed students to generate so-called 'saliency heatmaps' for a sample of user reviews. The saliency heatmaps highlight specific words or sets of words that most strongly contribute to the CNN sentiment predictions across the 300-dimensional word embeddings. For example, Figure 3 provides sample visualizations for four typical consumer reviews in the dataset. The purple shaded areas in the figure represent more salient terms for classification. In Figure 3, students can qualitatively see that in the first 2 examples, the algorithm has automatically learned from the context that the bi-gram "not working" or unigram "broken" most strongly contributed to the negative sentiment classification; whereas in the next 2 examples, terms like "it's working just fine" or "many thanks" or "wonderful charging station" strongly contributed to positive sentiment. These heatmaps help to somewhat demystify the inner architecture of the neural net algorithm, while offering the ability to check performance with relatively simple to understand visualizations. Next, we focus on additional visualizations for spatial analysis.



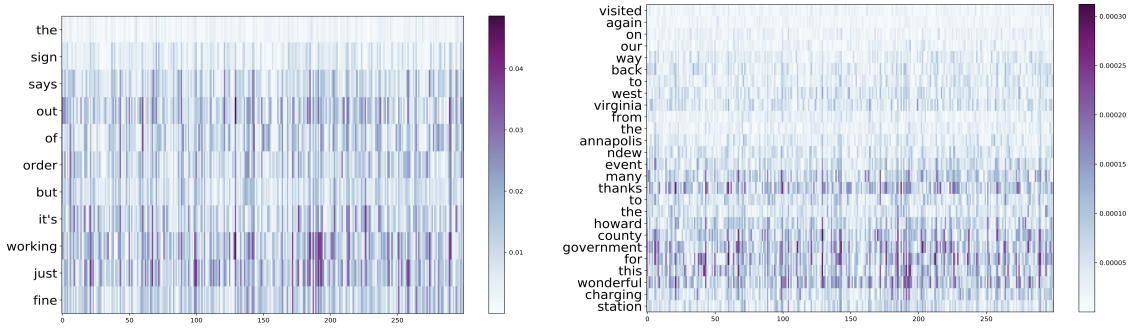


Fig 3. Saliency heatmaps. The top two heatmap examples correspond to negative sentiment predictions, while the bottom two heatmaps correspond to positive sentiment predictions. The words highlighted in purple in each of these heatmaps gives an indication of the most salient words that contribute to this model's classification output [16].

Visualization Tasks

For students in many social sciences fields, data visualization is often a critical competency that has not been well-integrated into the methods curriculum, particularly considering the wealth of newly available social and administrative datasets and related analysis techniques. For example, a few years ago the Georgia Tech Library initiated a campus survey asking what resources it should offer students as part of its effort to reimagine the future of library services. Data visualization services emerged as one of the top unmet needs across a variety of fields of study. Since its opening in Fall 2017, the Georgia Tech Library Data Visualization Lab has been supporting a variety of visualization tools for research and teaching. Previous cases in the literature have focused primarily on interactive visualization of large datasets primarily as a query processing problem [28]. However, given the efficient database support that now exists for scaling large database processing tasks with very little latency, we decided to focus this case on the visual data exploration. Here we describe the rationale for selecting user friendly cloud-based visualization tools to complement the machine learning analysis.

Use of Tableau Server: challenges and opportunities

The analysis of processed datasets such as from unstructured to structured data, due to the large sizes and complexity, often benefits from two key features. The first is online collaboration, which could be useful to engage students and managers in data literacy through web-based analytics. However, big data sources can also be proprietary and restricted from public access, a phenomenon that has been described as a big data divide [29], the separation that exists between those who curate and control data on a platform, and those who generate it. Consequently, a second important feature is the ability to balance data security issues with public accessibility. Given these two competing requirements, namely, online collaboration and data security, we selected a visualization tool for the ML output with both capabilities that could easily be implemented in a classroom setting and that is adaptable to distance or remote learning.

Data security

Following an independent IT procurement process not related to this research, we selected Tableau server as the platform. After considering various security requirements for Georgia tech provisioned software, we decided to keep the anonymized data on a secure cloud service maintained by Georgia Tech. We offered students the ability to connect to the source data with online visualization functions, but with restricted access to download or share the source data locally. This setting prevents direct sharing or distribution of raw data, and therefore, lowers the security risk. Tableau server also offers the option of hosting source data for non-restricted data sources on a cloud service with flexible security settings. According to their website, Tableau software is certified with the U.S. Department of Commerce in complying with the EU-U.S. Privacy Shield Framework (http://www.privacyshield.gov) in the event of any international personal data transfers. It is also in compliance with Sarbanes-Oxley requirements for financial data. It should be noted that, while students had no access to download the raw source data, once they completed and saved their visualizations on the server, students were able to download and share their own interactive visualizations. Although these restrictions on data accessibility are not normally required for public or open datasets, specifically for this case example, the source data is secured with instructor access only and 2-factor authentication.

Online collaboration

Another key objective was to introduce students to a cloud-based tool that offers the ability to create collaborative, interactive visualizations [30]. Modern visualization platforms have incorporated a number of interactive software features to a user's interface including zooming, panning, filtering, brushing, aggregation, feature or layer selection (examples include Esri ArcGIS, VIS, IVEE, Spotfire, Rivet, Polaris, Tableau). On Tableau server, once the workbook for a project is created, collaborators granted access can log in and work on the same visualization in real time either remotely or be collocated. This is a useful feature for distance or online learning situations in which students collaborate in small group or team settings. This approach can also readily be used in blended learning environments such as flipped classrooms, face-to-face courses, asynchronous distance learning or synchronous live learning [31]. After completion, the visualizations can be shared publicly including various interactive features. Cloud-based collaboration tools are often useful for big data projects that need different temporal or spatial analysis, multimodal interaction, or record linkages from a variety of databases [32]. Given the rapid pace of software enhancements, these features are subject further enhancements in the future.

Visualization Principles and Practice in Public Policy Studies

Basic design principles

During the visualization instruction, the data visualization librarian shared a set of best practices and design principles for the creation of interactive analytics dashboards to represent consumer sentiment. As part of this effort, a series of guided visualization tasks were demoed in the classroom as interactive tutorials. Students were encouraged to implement and design their

own visualizations using Tableau server. They were asked to build dashboards that incorporated five design best practices that we summarize below:

1. Select appropriate graphical elements, such as color, size and shape to help users identify information using culturally relevant norms [33];
2. Simplify cognitive load for end users by limiting the number of design elements [34];
3. Consider hierarchical graphical elements to showcase various levels of information [35][36];
4. Normalize scales and axes to represent large counts or data;
5. Add brief supporting text and phrases to connect the information flow from multiple frames into meaningful stories or dashboards.

Following these visualization best practices, a sample interactive dashboard was designed interactively to help users communicate spatial information from machine classified consumer sentiment data and evaluate infrastructure service provision. The sample analytics dashboard that we provided to students is shown in Figure 4 and an online interactive version is available at https://b.gatech.edu/2CJmFKf.

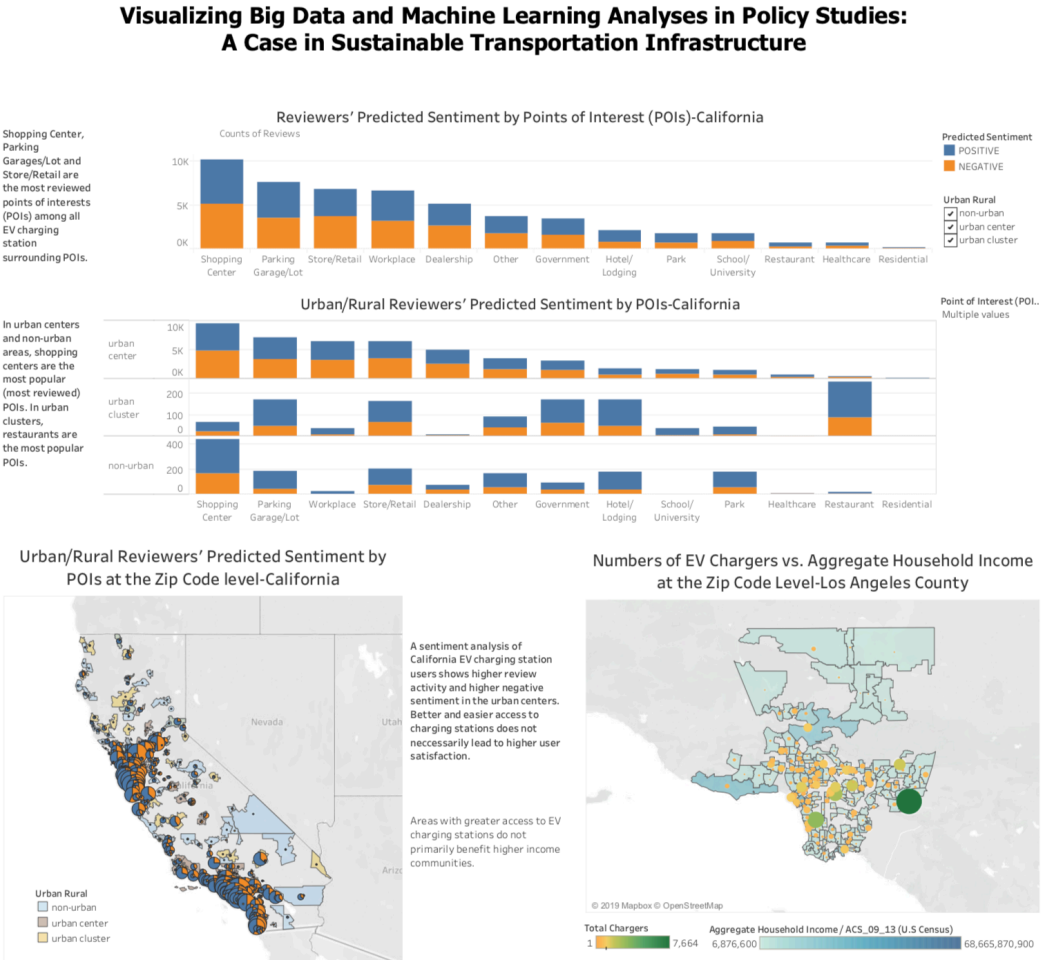


Fig 4. Sample Analytics Dashboard in Tableau

Generating Policy Insights

In this article, we highlight a policy teaching case in which derived data generated from text classification algorithms yields insights on a large amount of performance data, from unstructured text. Key to the interactive visualization tasks is the analysis of spatial features or dimensions. These guided visualization exercises have been designed to enable in-class discussions on a number of transportation policies related to vehicle electrification. This includes evaluation of consumer issues related to public EV charging services at various scales. For example, at a local scale, classroom discussions might focus on analysis of "EV capable" and "EV ready" policies that require property owners to reserve a certain number of spaces for EV charging as part of building codes and ordinances. At a regional scale, discussions might focus on calls for targeted policies to cultivate local EV clusters as an innovation priority for advancing the U.S. EV market [37], which benefits from spatial analysis. At a national scale, this case is also relevant for discussions about spillovers from Federal and State incentive policies such as rebates, income tax credits, subsidies, sales tax exemptions, and fee exemptions [38][39][40][41], all of which can benefit from spatial visualization tasks in both urban and non-urban environments.

By using a hands-on approach to data visualization for consumer analysis, students demystify the predictions or outputs of typical machine learning models. This helps to overcome the issue of 'opacity' between mathematical procedures in machine learning and human styles of semantic interpretation [42]. In a series of guided sessions, students are able to process streaming text data with relatively little technical background. In generating policy insights, first, students learn that easy access to a high density of EV charging infrastructure does not necessarily lead to higher consumer sentiment among EV users [15][16][17]. Second, urban centers generally receive more negative reviews at EV charging stations as compared to non-urban areas per spatial unit. Third, user experiences at EV charging stations also vary greatly by point of interest. For example, EV users in urban areas are most positive about charging stations located near restaurants, shopping centers and hotels/lodging destinations, while users in smaller urban clusters give the most positive feedback to charging stations located in residential areas and restaurants, with lower consumer sentiment observed in car rental locations and car dealerships. This heterogeneity suggests station hosts or operators may have different underlying incentives or ability to provide a high quality charging experience. Importantly, a regional analysis of EV charging stations in Los Angeles county, one of the largest areas in the U.S. for EV adoption and use, reveals that lower income neighborhoods are not necessarily disadvantaged in access to EV charging stations. Further analyses and visualizations could be done in different areas or geographies to examine these trends nationally.

We close with practical student feedback on the learning experience: "Social science students need more accessible classes that can bridge social science fields to cutting-edge techniques. This class really worked… it provided a wide range of knowledge in the field and helped me to have overall understanding about this [data science] field."

Case Study Questions

On Machine Classification and Properties of the Classifier

1. Describe the purpose of using a machine learning classifier and describe what information is typically needed to build confidence in sentiment predictions when comparing machine versus human classifications. Under what circumstances would a machine classifier be more effective than a human classifier for research evaluation? Be sure to consider a range of performance metrics.

2. Describe the classification accuracy of the baseline models (SVM, Logistic Regression) versus the specific CNN neural network based classifier. Do the CNN models consistently outperform the baseline models? Hint: Consider possible differences of means and distributions for 100 runs, and make an informed judgment based on the evidence.

3. Compare and contrast any possible tradeoffs in accuracy improvement versus computational time for CNN versus the non-neural net-based reference models. In what situations would one be preferred to the other?

4. What might be some ethical concerns raised regarding the uses of training data in this domain? How would you propose to mitigate them? Hint: Consider issues such as balance in the training data, nature of human raters, spatial sampling, incidental disclosure, reidentification, privacy and data security.

On Visualizing Classifier Performance

5. Generate saliency heat maps for the given pool of sample reviews and interpret the CNN model by giving context to the word embeddings. How well does a classifier identify the strongest words that contribute to the sentiment?

6. How would you ensure that your visualization framework remains robust to streaming (increasing) or anomalous data?

On Social and Policy Implications

7. Evaluate whether public investments in EV charging infrastructure have resulted in access disparities by income or other under-represented communities.

8. Determine whether urban on non-urban areas received the highest negative sentiment by consumer areas. What behavioral or policy theories predict your results? Do the results match your expectations?

9. Recommend a set of policies or incentives to address potential gaps, access disparities or issues in reliable service provision based on the evidence gleaned from the data.

Author Contributions

O.I.A. conceived and developed case study materials. X.M. developed visualization resources and evaluated software. S.D. validated case materials and replication code. All authors contributed to the original draft preparation, reviewing and editing of this paper.

Acknowledgments

We thank Sooji Ha for valuable classroom research assistance in the development of this case study.

Funding

We gratefully acknowledge funding support by the National Science Foundation (NSF) Award No. 1931980 and Award No. 1945332, Microsoft Azure for Research, and the Ivan Allen College Dean's SGR-C award.

Competing interests

The authors declare no competing interests.

Data Accessibility Statement

We have made the accompanying data visualizations publicly available at https://b.gatech.edu/2CJmFKf . Viewers are welcome to interact with the visualizations based on the anonymized, aggregated dataset. Due to privacy restrictions, the raw data cannot be posted publicly. However, open-sourced alternatives such as an open API for global EV charging infrastructure can be accessed for teaching purposes from OpenChargeMap. This is available online at the link in [43]. Instructors interested in accessing anonymized dataset supporting the findings of this case study or in jointly developing additional materials based on this work, may contact the corresponding author for data access. Additionally, we have also open-sourced our weights for the trained CNN model at the link in [44].

Code Availability

All replication code and instructions are available here.

References:

1. Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big Data Applications in the Government Sector. Communications of the ACM. 57(3): 78-85.
2. Jarmin, R. S., & O'Hara, A. B. (2016). Big data and the transformation of public policy analysis. *Journal of Public Policy Analysis and Management*, 35(3), 715-721.
3. Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015) Prediction policy problems. American Economic Review, 105(5), 491-495.
4. Pirog, M. A. (2014). Data will drive innovation in public policy and management research. *Journal of Policy Analysis and Management*, *33*(2), 537–543.
5. Boyd, D. & Crawford, K. (2012). Critical questions for big data, *Information, Communication & Society*, 15(5), 662-679.
6. Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, 76(6), 928-937.
7. Manoharan, A. P., & McQuiston, J. (2016). Technology and pedagogy: information technology competencies in public administration and public policy programs. *Journal of Public Affairs Education*, 22(2), 175-186.
8. Lane, J. (2016). Big data for public policy: the quadruple helix. *Journal of Public Policy Analysis and Management*, 35(3), 708-715.
9. Burns, W. (2017). The case for case studies in confronting environmental issues. *Case Studies in the Environment September 2017*.
10. Freeman, S., Eddy, S. L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111 (23), 8410-8415.
11. Isett, K. R., & Hicks, D. M. (2018). Providing public servants what they need: revealing the "unseen" through data visualization. *Public Administration Review*, 78(3), 479-485.
12. Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.
13. Athey, S. (2019). The impact of machine learning on economics. In, A. Agrawal, J. Gans, A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda*, Chicago, IL, University of Chicago Press, forthcoming.
14. Downie, D., & Bernstein, J. (2019). Case studies in the environment: an analysis of author, editor, and case characteristics. *Case Studies in the Environment*, DOI 10.1525/cse.2019.eic.001
15. Asensio, O.I., Alvarez, K., Dror, A., Wenzel, E., Hollauer, C., & Ha, S. (2020) Real-time data from mobile platforms to evaluate sustainable transportation infrastructure. *Nature Sustainability*, DOI 10.1038/s41893-020-0533-6
16. Alvarez, K., Dror, A., Wenzel, E., & Asensio, O.I. (2019). Evaluating electric vehicle user mobility data using neural network-based language models. *Transportation Research Board annual meeting*, Current Issues in Alternative Transportation Fuels and Technologies, Washington, D.C., 19-05863.

17. Ha, S., Marchetto, D., Burke, M.E., & Asensio, O.I. (2020). Detecting behavioral failures in emerging electric vehicle infrastructure using supervised text classification algorithms. Transportation Research Board annual meeting, Current Issues in Alternative Transportation Fuels and Technologies, Washington, D.C., 20-03461.
18. Scott, T.A., Ulibarri, N., Scott, R.P., Stakeholder involvement in collaborative regulatory processes: Using automated coding to track attendance and actions, *Regulation and Governance,* 2018.
19. Abernethy, J., Chojnacki, A., Farahi, A., Schwartz, E. & Webb, J. Active remediation: The search for lead pipes in Flint, Michigan. In [Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery #38; Data Mining, KDD 2018, 5–14 (ACM, New York, NY, USA, 2018).
20. Asensio, O. I. (2019). Correcting consumer misperception. *Nature Energy*, 4: 823-824
21. Asensio, O. I., & Delmas, M. A. (2017). The effectiveness of US energy efficiency building labels. *Nature Energy*, 2: 17033.
22. National Academy of Sciences, Engineering, and Medicine. (2018). Data science for undergraduates: opportunities and options, Washington, DC, National Academies Press https://doi.org/10.17226/25104
23. Sikder, Sujit & Herold, Hendrik & Meinel, Gotthard & Lorenzen-Zabel, Axel. (2019). Blessings of open data and technology: e-learning examples on land use monitoring and e-mobility. DOI 10.3217/978-3-85125-668-0.
24. Hopkins, D.J., King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54, 1, 229-247.
25. Corbett-Davies, S., Pierson E., Feller A., Goel S., Huq A., Algorithmic decision making and the cost of fairness, *KDD, 2017*.
26. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Distributed Representations of Words and Phrases and their Compositionality, *Neural Information Processing Systems, 2013*.
27. Li, J., Chen, X., Hovy, E., Jurafsky, D. Visualizing and understanding neural models in nlp. *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 681–691 (2016)*.
28. Godfrey, P., Gryz, J., &, Lasek, P. (2016). Interactive visualization of large data sets. *IEEE Transactions on Knowledge and Data Engineering, 28(8), 2142-2157*.
29. Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, 8, 1673-1689.
30. Martinez-Maldonado, R., Kay, J., Buckingham-Shum, S., & Yacef, K. (2019). Collocated collaboration analytics: Principles and dilemmas for mining multimodal interaction data. *Human-Computer Interaction*, 34(1), 1-50.
31. Hung H-C, Liu I-F, Liang C-T, Su Y-S. Applying Educational Data Mining to Explore Students' Learning Patterns in the Flipped Learning Approach for Coding Education. Symmetry. 2020; 12(2):213.
32. Liu, J., Tang, T., Wang, W. Xu, B., Kong, X., & Xia, F. (2018). A survey of scholarly data visualization. *IEEE Access*, 6, 19205-19221.

33. Nowell, L. T. (1997). *Graphical encoding for information visualization: using icon color, shape, and size to convey nominal and quantitative data* (Doctoral dissertation, Virginia Tech).
34. Healey, C. G. (1996). Choosing effective colours for data visualization. *Proceedings of Seventh Annual IEEE Visualization '96*, San Francisco, IEEE, 263-270.
35. Ware, C. (2012). *Information visualization: perception for design*. Waltham, MA: Elsevier.
36. Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6), 1139-1148.
37. Gordon, D., Sperling, D., & Livingston, D. (2012). Policy priorities for advancing the U.S. electric vehicle market. *The Carnegie Papers*, Washington, DC. Retrieved from http://carnegieendowment.org/files/electric_vehicles.pdf
38. DeShazo, J. R. (2016). Improving incentives for clean vehicle purchases in the United States: Challenges and opportunities. *Review of Environmental Economics and Policy*, 10(1), 149-165.
39. Holland, S. P., Mansur, E. T., Muller, N. Z., Yates, A. J. (2016). Are there environmental benefits from driving electric vehicles? The importance of local factors. *American Economic Review*, 106(12), 3700-3729.
40. Zambrano-Gutierrez, J. C., Nicholson-Crotty, S., Carley, S., & Siddiki, S. (2018). The role of public policy in technology diffusion: the case of plug-in electric vehicles. *Environmental Science & Technology*, 52(19), 10914-10922.
41. Sheldon, T. L., DeShazo, J. R., & Carson, R. T. (2019). Demand for green refueling infrastructure. *Environmental and Resource Economics*, https://doi.org/10.1007/s10640-018-00312-9.
42. Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*, 3(1), 1-12.
43. Open Charge Map - The global public registry of electric vehicle charging locations, available online at http://www.openchargemap.org.
44. Asensio, O.I., Ha, S. (2020). Trained CNN and LSTM model on EV charging station reviews for sentiment analysis. Figshare. https://doi.org/10.6084/m9.figshare.12044670

Figures and Figure Legends

Figure 1. **A general machine learning process flow.** After data collection and pre-processing, the data is split for training, testing and model validation.

Figure 2. **A Single Layer Neural Network Architecture for Sentiment Analysis.** The architecture of the Convolutional Neural Network (CNN) used for the experiment including the vector representations of words and the CNN structure as in ref. [15].

Figure 3. **Saliency Heatmaps.** In this figure, we compare four different saliency heatmaps with positive and negative sentiments, drawing student attention to specific words in the examples that strongly contributed to the classification output. The top two heatmap examples correspond to negative sentiment predictions, while the bottom two heatmaps correspond to positive sentiment predictions. The words highlighted in purple in each of these heatmaps gives an indication of the most salient words that contribute to this model's classification output.

Figure 4. **Sample Analytics Dashboard in Tableau.** The dashboard provides model visualizations for spatial analysis of consumer sentiment. They are available as interactive visualizations at: https://b.gatech.edu/2CJmFKf.