nature sustainability

ANALYSIS

https://doi.org/10.1038/s41893-020-0533-6

Check for updates

Real-time data from mobile platforms to evaluate sustainable transportation infrastructure

Omar Isaac Asensio[1]✉, Kevin Alvarez[2], Arielle Dror[3], Emerson Wenzel[4], Catharina Hollauer[5] and Sooji Ha[6]

By displacing gasoline and diesel fuels, electric cars and fleets reduce emissions from the transportation sector, thus offering important public health benefits. However, public confidence in the reliability of charging infrastructure remains a fundamental barrier to adoption. Using large-scale social data and machine-learning based on 12,720 electric vehicle (EV) charging stations, we provide national evidence on how well the existing charging infrastructure is serving the needs of the rapidly expanding population of EV drivers in 651 core-based statistical areas in the United States. We deploy supervised machine-learning algorithms to automatically classify unstructured text reviews generated by EV users. Extracting behavioural insights at a population scale has been challenging given that streaming data can be costly to hand classify. Using computational approaches, we reduce processing times for research evaluation from weeks of human processing to just minutes of computation. Contrary to theoretical predictions, we find that stations at private charging locations do not outperform public charging locations provided by the government. Overall, nearly half of drivers who use mobility applications have faced negative experiences at EV charging stations in the early growth years of public charging infrastructure, a problem that needs to be fixed as the market for electrified and sustainable transportation expands.

Global investment in electric vehicle (EV) charging infrastructure is estimated to reach US$80 billion by 2025[1]. In the United States, this investment growth marks an expected transition in policy support at the federal level to more aggressive actions at the state and local levels. The transportation sector is now the dominant source of CO_2 emissions in the United States[2]. By displacing gasoline and diesel fuels, vehicle electrification strategies have captured the attention of policymakers and analysts due to the expected public health benefits associated with reduced air pollution and tailpipe emissions[3–6]. However, while current EV infrastructure policies have focused on increasing the quantity of charging stations to meet future growth[7], not much attention has been paid to the quality of charging services, particularly at the consumer level. Service reliability is a key risk in the public provision of EV charging services and hence a critical barrier to large-scale technology adoption.

Some scholars contend that the private sector, under the right incentives, can more effectively deliver public fast-charging services as needed. Other scholars argue that large public investments in fast-charging infrastructure could crowd out private investments and lead to wasteful spending on charging locations that would have been built anyway (what economists refer to as inframarginal participation). Still other scholars argue that public charging serves a public good, particularly if sufficient incentives do not exist for private entrepreneurs and organizations to invest locally. This debate on public versus private provision of environmental public goods and services has a long tradition in economics[8,9] and public management[10,11], with mixed empirical evidence on whether decentralized local provision is more effective.

Subjective perceptions about the quality and reliability of public charging infrastructure are critical to building range confidence among existing EV owners[12–14]. Importantly, popular sentiment about EV charging station experiences could be even more critical to potential buyers in the EV purchase decision, particularly for consumers in underserved communities.

A major challenge to evaluating whether the current EV charging infrastructure is meeting the needs of the public is in access to available monitoring data[15]. This is because EV mobility data are largely user generated and are often owned by private entities[16,17]. For example, in the United States, charging transaction records are typically managed by tens of thousands of individual station hosts—each with the ability to independently set prices and charging policies (subject to State rules)—with no central repository or reporting requirements across network providers. As a result, given these high monitoring costs, national evidence on the quality of service provision in EV infrastructure has been scant.

In this article, we analyse evidence of EV charging station experiences in both public and private spaces and at major points of interest. We use machine intelligence to automatically classify user reviews in 651 core-based statistical areas (CBSAs) in the United States (Fig. 1). In doing so, we demonstrate the potential to use machine-learning to substantially reduce data aggregation costs by automatically classifying unstructured user reviews into positive and negative station experiences as an indicator of performance. On the basis of market data from 2011 to 2015, we show how a convolutional neural network (CNN) trained on large-scale social data learns domain-specific terms and, in effect, approaches the accuracy of human experts for sentiment classification. We then use machine classification as an input for econometric analyses that statistically adjust and mitigate potential observational biases in large-scale consumer data. We use this approach to evaluate

[1]School of Public Policy and Institute for Data Engineering and Science (IDEaS), Georgia Institute of Technology, Atlanta, GA, USA. [2]Department of Computer Science, North Carolina State University, Raleigh, NC, USA. [3]Department of Statistical and Data Sciences and Department of Government, Smith College, Northampton, MA, USA. [4]Department of Computer Science, Tufts University, Medford, MA, USA. [5]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [6]School of Civil and Environmental Engineering and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: asensio@gatech.edu

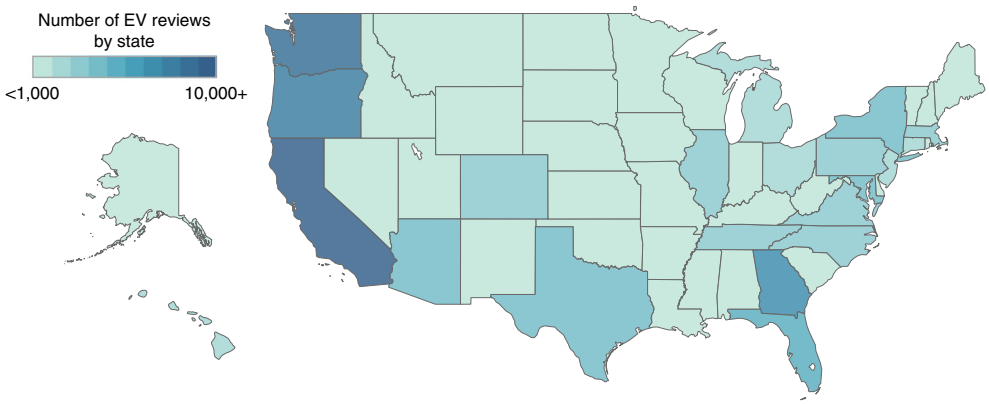


Fig. 1 | US map of active charge station reviews. This map shows the counts of EV charging station reviews per state from 2011 to 2015. The map was generated using ggplot/R (Source Data Fig. 1).

consumer sentiment and test hypotheses about service provision on a national scale.

We discuss performance in the context of sustainable transportation policies related to EV infrastructure. We further discuss directions for the use of consumer data and machine-learning tools in the analysis of government service delivery in near-real time and with dynamically growing datasets.

Mobility data

Mobile applications (apps) are changing the scale and techniques by which user behavioural data can be aggregated[16,18,19]. Digital platforms in mobile phones enable users to search, locate and pay for transportation services in real time. Given the rise in smart-phone use for transportation services, it is possible to analyse—subject to the necessary privacy protections—mobility decisions for large populations with digital infrastructure[20,21]. In the context of EVs, charging station locator apps help lower information and transaction costs. Users can search for available EV charging stations, pay for charging sessions and interact with other users by uploading station photos and writing station reviews for the EV community.

In this article, we analyse unstructured consumer reviews at 12,720 US charging station locations as provided by a popular EV charge station locator app. The data consist of 127,257 reviews from an estimated 25,133 registered and unregistered EV drivers during the period from 2011 to 2015. This includes a nationally representative sample of the US EV market with data aggregated from ten major EV charging networks in the United States.

Given the dynamically growing data size, it would be too costly for researchers or government analysts to hand classify these reviews for performance assessment. For example, at a rate of 100 reviews per hour, it would take a human expert about 32 work weeks to analyse reviews by hand. As a solution to this problem, we deploy machine-learning algorithms to automatically process unstructured reviews with natural language processing. This approach allows us to reduce processing times for research evaluation from weeks of human processing to minutes of computation.

CNNs. Recently, different types of neural networks have seen success in sentiment classification tasks for text data[22–25]. For example, CNNs first gained popularity in computer vision and have recently been demonstrated to be effective in several natural-language processing tasks[26–29]. However, these algorithms need to be adapted and optimized for specific domains before they can be useful. For this study, we implement a CNN and build on a model architecture similar to that proposed by Kim[22]. We choose this approach as CNN-based classifiers have been shown to achieve state-of-the-art results for sentence-level classification of short user-generated texts. This approach has an added benefit of automatically learning domain-specific semantics with lower dimensional representations versus existing approaches.

A key innovation of the CNN architecture is that it flexibly allows for unsupervised learning from pretrained word vectors while also allowing for supervised learning of domain-specific terms through back propagation[22]. In our implementation of deep-learning algorithms for EV mobility, we use pretrained word2vec word embeddings, which have been trained on approximately 100 billion words and phrases from Google news[30]. To capture domain-specific semantics, word embeddings can be updated as the model is trained. A summary of the key features of the CNN architecture, which includes the input word vector representations and the CNN procedure, is provided in Fig. 2. Additional implementation details and classifier procedures are provided in Methods.

Results

Machine classification. Using a CNN, we classify EV charging station experiences over a 4 yr period of rapid EV infrastructure growth from 2011 to 2015. We ask: how well do the machine predictions agree with human predictions? We know from our Cohen's $\kappa = .84$ achieved when building our training set that interrater agreement between human experts is high, but it is not perfect. As such, binary sentiment classification in this domain is difficult, even for human experts. With this in mind, it is encouraging that the CNN classifier achieved a sentiment prediction accuracy of 84.7% when compared with human labels (Table 1). To further demonstrate the efficiency of our classifier, we also report precision, recall and F1 measures of 0.86, 0.86 and 0.86, respectively. The use of deep-learning algorithms has recently been applied to short user-generated texts[22,25]. Here we demonstrate state-of-the-art performance to classify user reviews in the context of sustainable transportation and electric mobility.

We compared the performance of our CNN classifier to two non-neural net-based models, support vector machines (SVM) and logistic regression (LR), using the classic bag-of-*N*-grams approach. We also benchmark the performance of a CNN classifier versus some alternative neural net-based architectures, including variants of recurrent neural networks such as the well-known long short-term memory (LSTM) classifier[31,32]. To standardize the comparisons, we used the same word2vec word embeddings, with a single-layer (single directional) architecture for the LSTM classifier reported in Table 1. The LSTM classifier also achieved a respectable 83.1% in accuracy, with precision, recall and F1 measures of 0.85, 0.84 and 0.85, respectively. In additional experiments, we tested a variety of recurrent neural networks, including single and

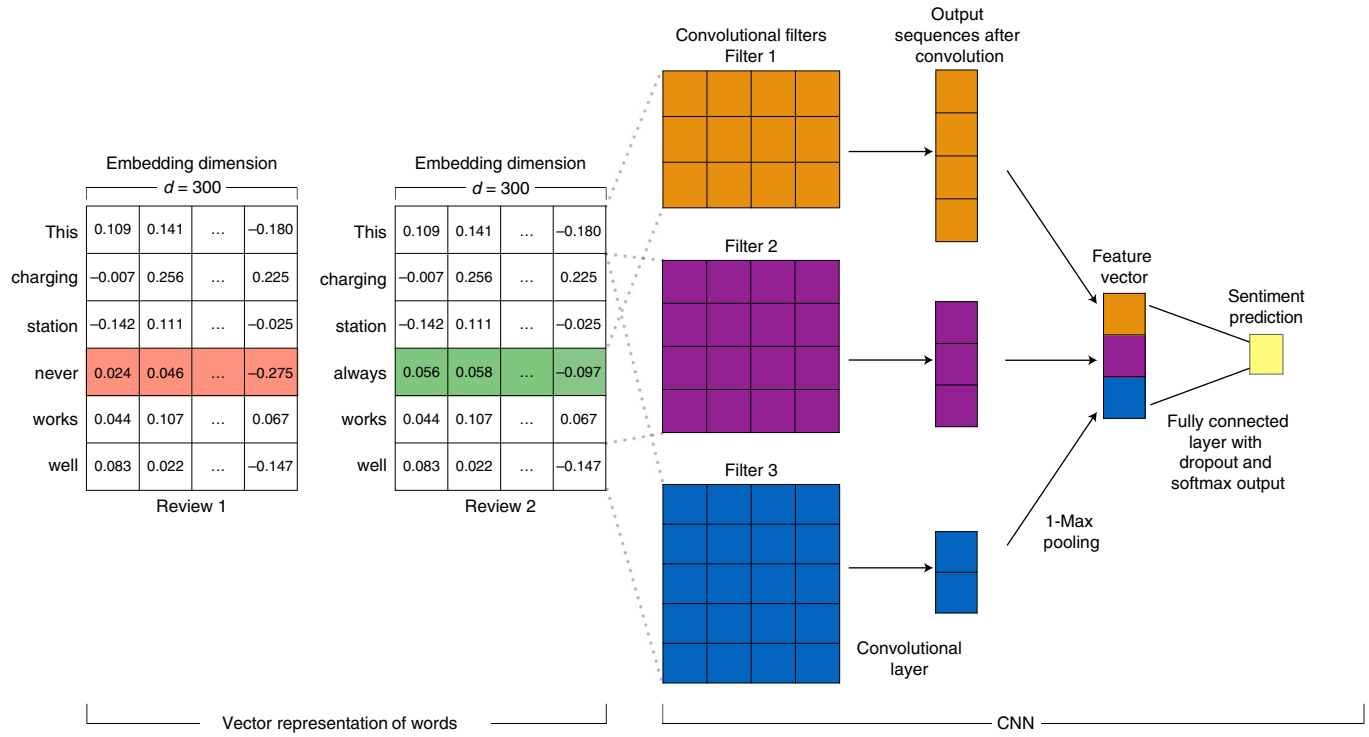


Fig. 2 | Model architecture for the CNN. The first stage shown depicts the vector representation of words in the review text with embedding dimension of 300. Each row in the embedding matrix is a vector representation of a word that captures information about word similarity. The second stage is the convolutional neural network, including a single convolutional layer for feature extraction. The convolutional filters of various dimensions learn which words to look for to predict sentiment. The extracted information from the convolutional layer is concatenated to make a feature vector that is processed to make a binary prediction.

Table 1 | Model performance

Model	Accuracy (%) (s.d.)	Precision	Recall	F1 score
CNN	84.7 (0.8)	0.86	0.86	0.86
LSTM	83.1 (0.9)	0.85	0.84	0.85
SVM	78.2 (0.8)	0.80	0.80	0.80
LR	79.1 (0.8)	0.80	0.82	0.81

Comparison of a single-layer CNN and an LSTM classifier versus other non-deep-learning baseline models: SVM and LR. For both of the baseline models, we use a bag-of-*N*-grams document representation with identical features. Results are reported on average values of 100 runs.

bi-directional gated recurrent unit and other LSTM networks with deeper architectures. These alternative neural nets yielded test accuracies in the 82–84% range, which also substantially outperform the SVM and LR baseline models not based on neural nets (Table 1). We find that for a similar level of performance, the LSTM models require about 50% more computing time (~41 min) versus the CNN-based models (~27 min) as clocked using a 16 GB memory allocation, which simulates an ordinary consumer laptop. Hence, for balanced training data, we find that CNN might be a preferred architecture for real-time analysis and implementation with streaming data (see Supplementary Discussion).

Domain-specific learning. In our series of experiments, we find that the CNN model identifies domain-specific patterns of natural language. For example, a commonly used term that may be recognized by subject matter experts, but not necessarily by the general population, is the notion of 'ICE-ing'. To be 'iced' or 'ICE'd' is an informal term that refers to cases in which an internal combustion engine vehicle is parked in a space normally reserved for EV drivers. ICE-ing is a common source of charge rage—the feeling drivers get when they are unable to find a charger. Its use reflects negative sentiment as it represents a violation of a community norm. For example: 'Came here on a Sunday around 11:30am and every spot was ICEd' or 'I was iced by a blue Dodge Journey'. For non-experts, these reviews might lead to ambiguous classifications due to insufficient domain knowledge otherwise common to EV drivers.

As neural network models are often criticized for their black-box nature, in Fig. 3 we provide sample visualizations of the salience of review text across the encoded word embeddings, using recently published protocols provided in refs. [33,34]. A higher saliency score assigned to a 300-dimensional word embedding indicates sensitivity to the final sentiment classification[33]. Following our example of highly contextualized terminology, Fig. 3 shows that 'iced' was the most salient term across the word embeddings, meaning that the algorithm has automatically learned the importance of this term. That artificial intelligence can detect 'ICE-ing' in this context and reach the accuracy of human experts, albeit in a matter of minutes of computation, is exemplary.

With this illustrative example, we show how artificial intelligence can be deployed to detect natural language associated with complex behavioural norms such as charging etiquette and other informal rules among a community of users. Such capabilities could also substantially reduce infrastructure evaluation costs and help equip utility managers and station operators with rapid response capabilities to improve service times. We suggest future research to explore further uses of machine intelligence to identify behavioural mechanisms related to charge rage, congestion and other station failures. In the next section, we use our best prediction model to test common assumptions about charging behaviour in public and private spaces and at key points of interest.

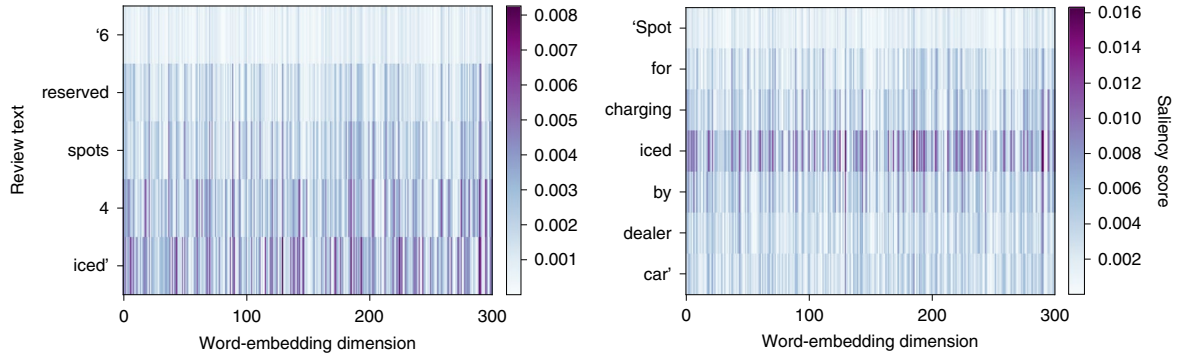


Fig. 3 | Saliency heatmap for reviews with a domain-specific term 'iced'. The visualization shows a higher saliency score for the term 'iced', demonstrating that the CNN algorithm successfully learned the importance of the term, as recognized by human experts (Source Data Fig. 3).

Sentiment analysis. We find evidence of widespread EV charging station use in all major US geographic areas (Figure 1). To compare the incidence of negative sentiment for econometric analyses, we created a negativity index of conditional probabilities across stations, where 0 means all reviews at a given station location are positive, and 1 means all reviews at a given station location are negative (equation (1)). A higher predicted sentiment probability (closer to 1) would therefore not be desirable.

The mean predicted sentiment across all station reviews in both urban and non-urban areas is 0.44. While this number might not seem high at first, it is analogous to predicting a negative experience four out of ten times that a driver goes to a gas station to fill up a car and writes about the experience. Further, if we aggregate the sentiment score by user ID per year, this is nearly half of all drivers facing negative experiences. By contrast, in a recent paper on consumer reviews from smart mobility apps, gas station fuelling service and real-time data are primarily classified as positive experiences by consumers in a similar use[35]. Indeed, sentiment analysis from mobile navigation apps finds that only 16% (9 out of 55) of service-related keywords are negative (for example, location sharing, processing speed and arrival time estimation)[35]. We argue that a greater focus on the quality of the EV charging experience is needed. For additional examples of sentiment analysis in other contexts such as public opinion polling, environmental impact statements, measuring cultural norms and network effects, see refs. [35–39].

Discussion

We discuss the results for public versus private stations, urban versus rural, and by points of interest.

Public versus private stations. Theory predicts that under the right incentives, private charging stations should outperform those run by government entities[9]. However, in practice, it is unclear whether sufficient incentives exist for private station hosts to maintain a high level of service quality, especially in the reselling of electric power, where capital cost recovery is often challenging and retail electricity prices are low. Here we test the hypothesis that private charging stations more effectively deliver charging services versus public stations provided by the government. We considered a broad definition of public stations such as those that have been geolocated at points of interest (POIs) that include government and municipal buildings, public libraries, rest areas, transit centres, public parks and visitor centres. We define private stations as those that have been geolocated at POIs that include hotels, retail/food establishments, shopping centres, healthcare facilities, workplaces and other non-residential locations. As in many contexts related to the private provisioning of public goods, we might expect to see evidence that charging stations at privately managed locations outperform those that are publicly owned or managed.

In Supplementary Table 1 (Extended Data Table 1), we provide descriptive statistics for the raw counts of machine-classified reviews at both public and private charging destinations. Contrary to expectation, we do not find a statistically significant difference in the mean predicted sentiment between public and private charging station locations nationally (see Fig. 4). We validated this finding by adjusting for factors driving selection to review and other observable location characteristics. In addition, we considered a more-narrow definition of public chargers with POI restricted to government-only stations and verified statistical parity in consumer sentiment between public and private stations (see Supplementary Table 2; Extended Data Table 2). Additional details on public and private location designations is provided in Methods.

We interpret this finding in two ways. First, our results indicate that private charging station locations do not outperform those that may be publicly owned or managed. Second, from a public choice perspective, our results provide some evidence that the private provisioning of EV charging services could be an alternative to large, publicly managed infrastructure. For example, one anonymous reviewer wrote about the substitutability of a public charger for a private charger: 'Be careful if you plan on charging here, there are two cars that tend to bogart these chargers try the city hall chargers'. Evidence of statistical parity in consumer perceptions between public and private charging locations addresses one concern raised by the National Research Council on barriers to EV infrastructure growth[40]. We caution, however, that our performance indicator captures popular sentiment from the standpoint of national consumer reviews, and not a power systems delivery perspective, which requires further investigation and integration with consumer data.

As shown in Fig. 4, we find that paid charging stations receive a higher proportion of negative reviews as compared with free stations. Not surprisingly, this result holds whether the station is in a public or private location. This finding suggests users may have higher expectations for service reliability when paying for charging services. It is plausible that EV station location, whether public or private, may not be the dominant factor affecting service reliability. For example, publicly owned stations could have enjoyed similar (or perhaps even higher) levels of operation and maintenance subscription services. In the next section, we use location microdata to investigate possible regional differences.

Urban versus rural areas. We compared the performance of stations in urban and non-urban areas. According to one view, EV drivers in areas with fewer charging stations are more likely to experience issues of range anxiety, possibly leading users in these areas to publish more negative reviews. Therefore, from a resource-availability hypothesis, areas with greater access to charging infrastructure should garner the most positive reviews. Interestingly, we do not

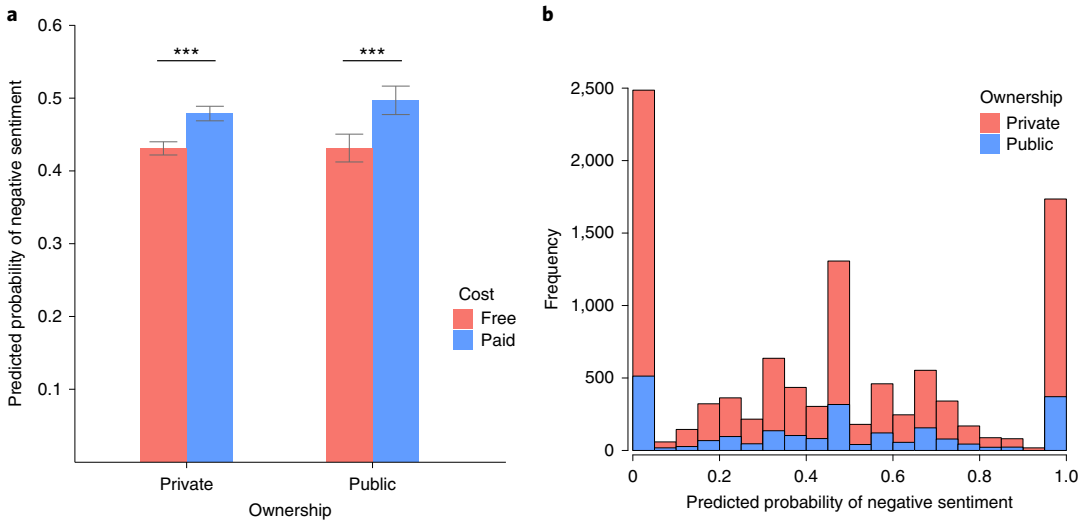


Fig. 4 | Predicted probability of negative sentiment in public and private spaces. a, Charging stations for pay have a higher predicted probability of negative sentiment in both public and private stations (***$P < 0.001$). We find no statistically significant difference in consumer sentiment between public and private stations ($P > 0.150$). The error bars indicate ±95% confidence intervals (Source Data Fig. 4). **b**, We find differences in the distribution of the sentiment scores (Kolmogorov–Smirnov test, ***$P < 0.001$).

find evidence for this hypothesis. The highest incidence of negative sentiment is not in rural areas or even smaller urban clusters, but rather, the dense urban centres (see Fig. 5a). This is intriguing because approximately 89% of all user reviews are in urban centres (Supplementary Table 3; Extended Data Table 3). After controlling for important station location and timing factors, we find that urban charging stations exhibit a statistically significant 12–15% increase in the predicted negativity score as compared with non-urban locations (Fig. 5a and Supplementary Table 2, Models IV–VI; Extended Data Table 2, Models IV–VI).

Our finding that EV charging stations in urban centres significantly underperform those in smaller urban clusters or rural areas where population and station densities are lower could be indicative of a broader range of service quality issues in the largest EV markets. For example, many users report a lack of functional stations upon arrival, as well as issues related to congestion or lack of availability: 'some person is just pulling plugs without any note; i'll review footage on my security cam' or 'Both spots taken. One by a Volt that's finished charging... Seriously time for more EVSE stations'. In Supplementary Tables 4 and 5 (Extended Data Tables 4 and 5), we summarize the predicted (negative) sentiment probabilities for both free and paid charging stations in the 18 largest US metro areas and top 20 US states by number of reviews. Although user reviews exist in all 50 states, the dominant source of activity is California, with 54,684 reviews, or 43% of all consumer reviews in the dataset. The Los Angeles metro area, for example, is the largest CBSA for charging station reviews through 2015, with 22,878 reviews, or 18% of all reviews in the dataset. The mean predicted negative sentiment in Los Angeles ranges from 0.52 to 0.59, which means a given user is more likely to report a negative consumer charging experience than to report a positive one. This is considerably higher than the estimated US national average sentiment score that we report of 0.44. These results indicate that service reliability is already a factor impacting consumer sentiment in the largest EV markets. Although quality is important in the consumer experience, we note that EV users are often confronted with fewer substitutable fuelling options, particularly in rural areas. Next, we evaluate the results by POI.

POIs. We summarize the results of our sentiment analysis by POI in Fig. 5b. The highest-performing private stations are at points of interest such as hotels/lodging destinations, restaurants and food establishments, and convenience stores. This is to be expected as private-station hosts in these locations often provide subsidized or complimentary EV charging services as a way to attract and cater to specific clientele. This suggests incentive-based management practices. The highest-performing POIs also include parks and recreation as well as RV parks and visitor centres. All of these POIs are associated with travel destinations. On the basis of our examination of reviews at these locations, we believe that destination range anxiety could factor into positive reviews at these locations since drivers may be willing to sacrifice challenging conditions for the needed charge. Some of the lowest-performing POIs include car rental locations and car dealerships. This is not inconsistent with recent evidence on car dealership practices at the point of sale, which have been documented as promulgating barriers to EV adoption[41]. Workplace and mixed-use residential locations with retail establishments are also relatively low-performing POIs. For example, many EV users at workplace and mixed-use residential locations complain that EV stations can be difficult to access or that there is poor signage for public accessibility. We provide detailed POI point estimates net of statistical controls in Supplementary Table 2 (Extended Data Table 2).

Policy implications

Large-scale data from digital platforms can offer benefits for research evaluation efforts. We show that using computational tools, it is possible to develop more sophisticated performance indicators from unstructured data that offer the potential to update in near real time. This is a major step forward from current practice that relies on indirect travel surveys or simulations, which can be costly and time intensive to administer[42]. We argue that consumer data should be prioritized when designing policies related to EV infrastructure access. This is particularly important in the design of 'EV ready' or 'EV capable' policies in building codes and ordinances that require new buildings, for example, to install and maintain a certain number of EV charging stations [43]. Such policies have grown in popularity in many cities such as Palo Alto, Denver and Atlanta, largely without data or deliberation about service quality from existing EV users or consumer groups.

Mobile apps can aggregate consumer data automatically at scale, but independent station hosts and operators currently have

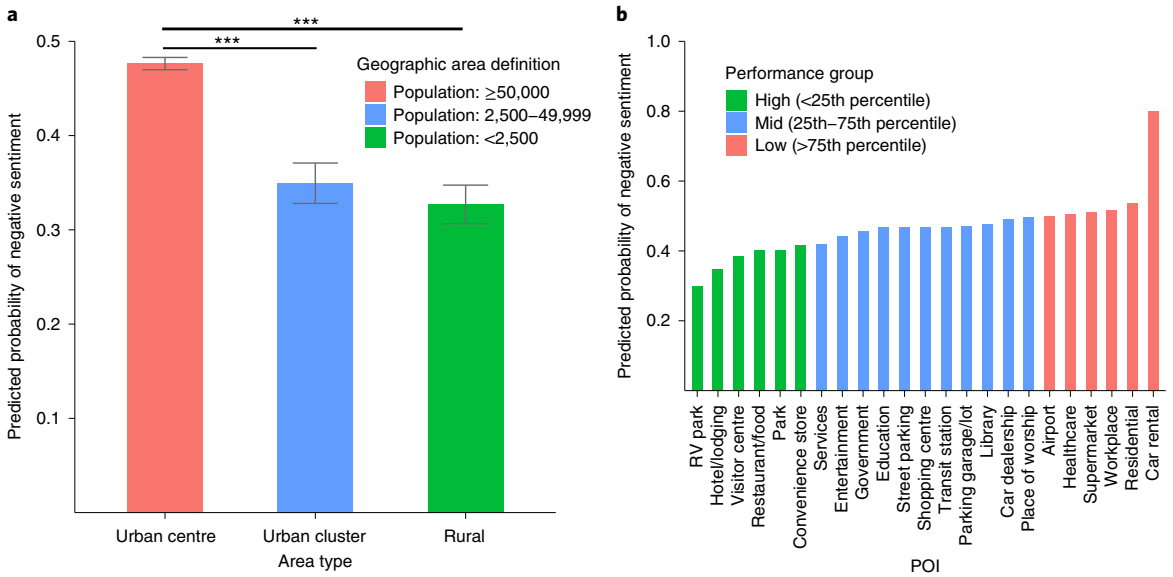


Fig. 5 | Predicted probability of negative sentiment by geographical area and POI. a, EV chargers in urban centres have significantly higher negative sentiment scores compared with smaller urban clusters and rural areas (***$P < 0.001$), net of statistical controls. The error bars indicate ±95% confidence intervals. **b**, The rank order of consumer sentiment by POI (Source Data Fig. 5).

little incentive to share data across network providers. Centralized reporting and secure data sharing across charging networks and utility jurisdictions would allow for more-efficient resource planning decisions, particularly in resilience considerations between power systems delivery and emerging transportation infrastructure. One key exception to platform data sharing is the listing of EV charging stations maintained on the Alternative Fueling Station Locator hosted by the US Department of Energy under the Clean Cities Program. While an invaluable tool, this digital repository of EV charging stations does not currently contain real-time availability or network status information. We argue that policies to encourage greater information sharing as well as standardization in the quality of charging service delivery are necessary.

In this article, we show how machine intelligence can approach the accuracy of human experts for sentiment classification tasks, while showing promise for automatically learning domain-specific terms in emerging EV communities. Machine-learning techniques can automate the process of discovering new mobility patterns and detecting behavioural failures from consumer data, but they do not replace the need for a human in the loop. Note that, because the classifier is trained by a human, the classifier is only as unbiased as the human rater. Not all consumer reviews can be relevant or actionable. Nevertheless, by expanding administrative records with real-time streaming data, it is now possible to track station performance in both accessible and remote areas in ways not previously possible. Further, the use of machine intelligence as a pre-processing step for policy analyses can be helpful to determine consumer requirements in both coverage and demand assessments related to transportation infrastructure[44]. This focus on big data and real-time mobility will become increasingly important as driver incentives and other supply-driven policies designed to reduce externalities from EVs do not typically address or affect driver behaviour[45,46].

As EV infrastructure grows, we argue that it is not only the quantity of available stations that matters to consumers, but also the quality of the charging experience. A key focus for quality improvement should be in the urban centres, where reports of ICE-ing and lack of available or functional stations are prominent and appear to drive negative consumer experiences. However, further research is necessary to determine the most important mechanisms of user dissatisfaction. Community interactions also reveal emerging norms about charger etiquette and prosocial behaviour primarily designed to help others in the community. We expect US$80 billion in new investment in EV supply equipment over the next few years[1]. On the basis of evidence from consumer data, we argue that it is not enough to just invest money into increasing the quantity of stations, it is also important to invest money into reliable infrastructure that actually works.

Methods

Overview of CNN architecture. The first stage of the CNN deep-learning architecture shown in Fig. 2 depicts the matrix representation of the review text. Similar words are closer together in the vector space than are disimilar words[23]. In our implementation, we used pretrained word2vec word embeddings, which have been trained on approximately 100 billion words and phrases from Google news[30]. To capture domain-specific semantics, word embeddings can be updated as the model is trained.

The second stage is the CNN. Convolutional filters learn what words to look for in the reviews. There can be several convolutional filters per filter size[47]. Filter heights in CNN models represent the dimension of the N-gram within the text. We used filter heights of 3, 4 or 5. For example, a filter size of 3 scans through the word embeddings as a bag of 3-grams in the process of convolution. For example, in the review text 'got a charging error first few times got AV to reset it remotely and it worked finally', the convolutional layer with filter size 3 results in the following 3-grams: 'got a charging', 'a charging error', 'charging error first', …, 'and it worked', 'it worked finally'. Then, the 1-max pooling procedure as described in ref. [29] selects the most-important 3-gram for the prediction task. The 1-max pooling outputs from multiple filters are then concatenated to form a feature vector. These features are passed onto a fully connected softmax layer with dropout regularization, whose output is the probability distribution over labels[22]. In the current example, the 'a charging error' 3-gram was the most predictive feature for negative sentiment classification.

Selection of CNN hyper-parameters. We used various strategies to select our CNN hyper-parameters. Building on previous literature, we selected 1-max pooling, dropout regularization[48] with a rate of .3 and a ReLU activation function in our convolutional layer, as these hyper-parameters have been shown to improve accuracy[47]. In particular, the dropout technique was implemented to prevent overfitting[48]. In our implementation, we confirm that an L2 constraint had no discernible performance improvement, and therefore we do not include it in our model[47]. Other hyper-parameters include a batch size of 128; learning rate of .001; filter heights of 3, 4, and 5; 100 filters for each filter height. Filter widths are 300, which are set to the dimensionality of the word embeddings.

LSTM implementation and hyper-parameters. We implemented a one-layer LSTM model to make it comparable to the CNN model, using protocols described in ref. [49]. We then performed basic optimization. For example, we used the same

word2vec word embeddings with 300 dimensions. We used the same train-test split and preprocessing as in the CNN model. In both LSTM and CNN models, we used the Adam optimizer, with a learning rate 0.001. Other key hyper-parameters include cell number 64, dropout rate 0.6 and recurrent dropout rate 0.2.

Curating the training data. In any supervised classification task, it is necessary to obtain ground-truth labels. To generate these labels, two research assistants served as human expert raters. We conducted several focus groups to decide on a common set of rules for classifying the training data. Each human rater independently coded an identical set of 8,953 reviews, which were chosen as a representative sample from the 127,257 reviews. We initially investigated the accuracy of a three-class model—positive, negative and neutral classifications. However, neutral classifications were found to be very difficult for human rater tasks in this domain. We therefore achieved the best interrater reliability ($\kappa = 0.84$, SE = 0.7)[50] by treating this as a two-class problem, meaning reviews reflect a binary sentiment only (positive/negative). Although imbalances in training data typically present potential problems for machine classification, in Supplementary Table 6 (Extended Data Table 6), we show that reviews in our training data are highly balanced in polarity.

As part of model validation procedures, the 8,953 hand-classified reviews were randomly split to 80% of training data and 20% of testing data. We also tried other random splits and cross-validation procedures and found quantitatively similar results.

Examples of consumer reviews. To provide additional details of the human labelling tasks for training, we provide some examples of both positive and negative labelled reviews.

Positive:

- 'Charged! When I called Blink CS before I traveled they said a tech had been here to fix this station, and I am happy to report it is!'
- 'Huge solar panels power this amazing station!!'
- 'Surprisingly not ICEd at 5:45pm on a Tuesday. Stall 2B seemed slow, delivering only 28 KW at 45% SOC. moved to 3A.'

Negative:

- 'OUT OF SERVICE AGAIN! This station is a waste of time'
- 'Never lucky enough to get a spot to charge, someone's always there. Good luck!'
- 'All spots full right now. Charging BMW i3 with SAE combo.'

Statistical uncertainty. Starting with a random data split and initialization, we report the statistical uncertainty of the test accuracy for the CNN classifier for 1,000 runs. The mean test accuracy is 84.6% (minimum 82.2%, maximum 86.9%) with s.d. 0.79 (see Supplementary Fig. 1; Extended Data Fig. 1).

Econometric analysis. Following machine classification, we conduct econometric analyses to mitigate possible observational biases and statistically adjust for station location and timing factors. Our unit of analysis for each review is at the station level.

Measuring outcomes of interest. For a given charging station location i and review period year, we define the negativity score as the count of negative reviews as a fraction of the total count of reviews:

$$\text{NegativityScore}_{i,\text{year}} = \frac{\text{Count of negative reviews}_{i,\text{year}}}{\text{Total count of reviews}_{i,\text{year}}} \quad (1)$$

By construction, the share of negative reviews at a charging station is normalized to lie in the unit interval [0, 1]. Boundary observations of the dependent variable at 0 indicate that all reviews at a charging station are positive, and boundary observations at 1 indicate that all reviews at a charging station are negative. A higher negativity score is undesirable. We also group the charging stations by location group, g (that is, there can be more than one station ID at a given location) and the year of the review to provide a rate of users leaving a review relative to the amount of use of the class of station. Users can check in to the platform and leave a review, which enters into the count of reviews, or check in without leaving a review, which enters into the count of other check-ins. We define the re view rate as:

$$\text{ReviewRate}_{i,g,\text{year}} = \frac{\text{Count of reviews}_{i,g,\text{year}}}{\text{Count of reviews}_{i,g,\text{year}} + \text{Check}-\text{ins}_{i,g,\text{year}}} \quad (2)$$

Fractional response models. We used the outputs of the CNN classifier as a pre-processing step for econometric analysis of consumer sentiment. Given the limitations of linear estimation methods such as ordinary least squares (OLS) for bounded dependent variables, we implemented a fractional response model (FRM) for the probability share data based on the quasi-maximum likelihood (QMLE) estimator[51,52]. We present some elements of FRM models as developed by Papke and Woolridge[51,52] and then apply them to machine-learning outputs in mobility data. In the standard FRM setup, we are interested in the conditional expectation of the fractional response variable $y_{i,t}$ on a group-specific vector of explanatory variables $\mathbf{x}_{i,t}$ such as:

$$E(y_{i,t}|\mathbf{x}_{i,t}) = G(\mathbf{x}_{i,t}\boldsymbol{\theta}), \quad i = 1, \ \ldots, \ N \quad (3)$$

where $G(\cdot)$ is a nonlinear function such as the cumulative distribution function that satisfies $0 \leq G(\cdot) \leq 1$, the fractional dependent variable is defined only on $0 \leq y_{i,t} \leq 1$, and $\boldsymbol{\theta}$ is a parameter vector of interest. Observations at the extremes of the outcome distributions are estimated directly using the Bernoulli log-likelihood function, given by:

$$l_{i,t}(\boldsymbol{\theta}) \equiv y_{i,t}\log[\mathrm{G}(\mathbf{x}_{i,t}\boldsymbol{\theta})] + (1 - \mathrm{y}_{i,t})\log[1 - \mathrm{G}(\mathbf{x}_{i,t}\boldsymbol{\theta})] \quad (4)$$

In our dataset, some charging station reviews may be classified as all negative or all positive at a given location. Given the presence of boundary observations at 0 and 1, the pooled Bernoulli quasi-maximum likelihood estimator of $\boldsymbol{\theta}$ does not require dichotimizations of the dependent variable and is computed as:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} l_{i,t}(\boldsymbol{\theta}) \quad (5)$$

We note that this approach overcomes three important limitations found in comparable methods. First, we account for the bounded nature of the data and do not assume a linear conditional mean or constant linear effects, which requires stronger assumptions for estimation. Second, commonly used log-odds methods are not well defined for boundary values 0 and 1 present in the data and often require ad hoc adjustments such as arbitrarily chosen constants. Third, methods based on two-limit Tobit models may be appropriate for censored data with boundary observations at both limits, but their application to fractional data that are not defined outside the boundary limits is hard to justify. For a more-detailed review of fractional regression models, see Ramalho et al.[53]

Selection effects of providing a review. The decision to provide a review is a voluntary one. It conditions the interpretation of information developed by analysing a sample of reviews. Charging station activity outside the digital platform is inherently unobservable. To address possible limitations due to selection effects, we attempt to explain the likelihood of giving a review as a function of characteristics of the charging location and timing. In equation (2), we normalize the empirical review counts by total platform engagements, including user check-ins without reviews. In this way, we are able to normalize our estimates of the importance of explanatory variables on the empirical review rate by a measure of total charging station usage beyond review activity. For example, during the period of study, there are 276,749 total user check-ins on the platform, of which 127,257 contain English-language reviews.

Our main estimating equation relates the review rate and negativity score as a function of one or more of the explanatory variables. This includes POI information, geographical area such as urban, suburban or rural, the type and count of networks available, the type and count of station connectors available, and our designation as public stations on the basis of station geolocation. Due to data limitations, we could not adjust for car type of the driver as that information is voluntary, so we had a biased subsample. In addition, we also tested specifications that included the proprietary station quality rating (numeric score 1–10) as a proxy for possible unobserved heterogeneity. We estimate the following general equation:

$$\begin{aligned}\text{Outcomes}_{i,g,\text{year}} = G\{&\alpha_{i,\text{year}} + \text{Public}_{i,g} + \text{POI}_{i,g} \\ &+ \text{Networks}_{i,g,\text{year}} + \text{Connectors}_{i,g} + \text{Rating}_{i,g}\}\end{aligned} \quad (6)$$

Observable location characteristics include the type and number of networks available (for example, Chargepoint, Blink, SemaConnect, Aerovironment, EVgo, Tesla Supercharger, GE Wattstation and so on), the type and number of connector plug technologies (for example, J-1772, CHAdeMO, SAE Combo, Tesla supercharger, NEMA plug and so on) and other driver-identifiable location attributes by POI. To mitigate possible unobserved heterogeneity, we also include the station rating as a proxy variable for unobserved quality attributes.

In Supplementary Table 2 (Extended Data Table 2), we report the main results. The main drivers of the review rate include geographical area (whether urban or non-urban) and point of interest location information. We also find a statistically significant effect of the type and number of station connectors available and the type of charging network such as the service provider, but not the number of networks available at a station location, which can range from one to three networks at a location ID. This suggests choice in charging network service provider is not yet a significant factor. Given our main interest in the public provision of charging services, we confirmed our finding of no significant effect of public locations (or more narrowly defined government-only locations) on the review rate. This result is robust to our proxy for unobserved quality attributes as measured by the station quality rating provided to us by the platform provider (Supplementary Table 2, Models II, III; Extended Data Table 2, Models II, III). Overall, for factors driving the selection to review, location matters, as do the network type, connector technology and other quality-related factors.

In Supplementary Table 2 (Extended Data Table 2), we do not show the point estimates for individual networks or plug types, but these results are available upon request from the corresponding author.

We conditioned on all observable characteristics from our aggregate selection analysis to then compute the average partial effects for factors driving the negativity score. The analysis reveals that urban chargers account for a statistically significant 12–15% increase in the negativity score compared with non-urban locations (Supplementary Table 2, Models IV–VI; Extended Data Table 2, Models IV–VI). Similarly, we also confirmed our finding of no statistically significant effect of private versus public stations, which is robust to both a broad and narrow definition of public stations and unobserved quality factors.

In Supplementary Table 7, we provide supplementary regression results comparing the performance of the FRM approach with a standard OLS estimator. While we find the estimates to be qualitatively similar, we see that FRM generates more-conservative estimates compared with OLS, which overestimates the magnitude of the effects, as expected.

Urban versus rural definitions. For spatial analysis, we merged the geocoded station location data with geographical designations using standard US Census definitions[54]. These include urbanized areas (populations greater than 50,000), smaller urban clusters (populations between 2,500 and 50,000) and rural areas (populations less than 2,500). According to the 2010 Census, there are 486 urbanized areas and 3,087 urban clusters in the United States. All designations contained reviews. For descriptive statistics of binary sentiment classifications by geographic area, see Supplementary Table 3 (Extended Data Table 3).

Comments on defining public and private stations. To determine whether the chargers on public properties were also publicly owned and managed, we contacted a random sample of 170 public EV charging locations, stratified by network (Blink, ChargePoint and so on). We then attempted to contact each location through a combination of email and phone calls to ask the following questions: Are the charging stations at this property owned by the organization? Are the charging stations at this property managed by the organization? We also contacted several major EV charging networks directly (for example, SemaConnect, ChargePoint, GreenLots and Blink) to determine whether they operate/maintain charging units on behalf of their customers. For example, we found that while GreenLots network manages all of their stations on behalf of station hosts, station owners from the other three major networks we contacted can decide whether they want to enter into a contract/warranty for servicing. Overall, we found four possibilities regarding station ownership and maintenance on public properties:

- Stations are both owned and managed by public entities (such as those in Colton, California).
- Stations are owned by public entities but managed by private EV charging networks (such as the one at the Anaheim Intermodal Transit Center in Anaheim, California).
- Stations are owned by public entities but managed by a local contractor (such as the station at Roswell City Hall in Roswell, Georgia).
- Stations are neither owned nor managed by public entities (such as the station at the Minnesota Department of Natural Resources in St. Paul, Minnesota).

After contacting 170 stations, we were able to obtain answers to our management question at 32 locations. Of these 32 locations, 10 were managed by the public entity, and 22 were managed by either an EV charging network or a private company. We were also able to get answers to our ownership question at 23 locations. Of these 23 locations, the stations at 14 locations were owned by the public entity, and the stations at 9 locations were not. We believe that the management structure can potentially be an important driver of proper functioning of EV chargers and, hence, the consumer experience. However, the managerial aspects of public versus private operation, while outside the scope of the current paper, we highlight as important differences for future research.

Study limitations. While we demonstrate gains using machine-learning in this domain, there remain key areas for technical improvement. First, it may be necessary to increase the size of the training data to achieve even higher convergence between human and machine classifications. This is especially relevant in dynamically growing social datasets where topic categories may be broad. For reference, we calculated an alternative agreement score between the human predictions and machine predictions by treating the machine as a separate rater. The resulting $\kappa = .68$ suggests additional optimization could be necessary to increase reliability scores. However, due to computational complexity, it may be difficult to fully optimize all hyper-parameters to reach a global optimum. Second, future work can explore deeper architectures and optimal filter sizes. For example, a recent paper on very deep CNNs for text classification reports optimal results with up to 29 convolutional layers[55]. In a sensitivity analysis of CNN, one approach proposed by Zhang and Wallace is to conduct 'a line search over the single filter region size to find the 'best' single region size'[47]. This could be a promising approach to further improve accuracy in subpopulations of review types or in training sets with different types of human raters. We leave this as future work.

In this paper, we implement recent deep-learning approaches to automatically learn text representations for sentiment analysis, but we do not demonstrate their performance for topic labelling, which could open new directions for discovery of behavioural failures. We leave this task for future work. We also point out that although text data are time stamped, it is in many cases not possible to directly observe the contemporaneous power systems delivery to validate consumer claims. To verify the operational status or other specific issues, consumer and power data must be linked in information systems, which is a major challenge. Another limitation of our analysis is that while we are able to quantitatively evaluate sentiment from consumer reviews, additional information is needed to identify the psychological basis for negative charging experiences. It would be useful to develop topic classifications and accompanying training data with ground-truth labels that describe the various sources of negative consumer experience. This might allow for deeper identification of mechanisms and algorithmic classifications for policy analysis.

Data availability

We provide the weights of the trained deep-learning models. These datasets generated and/or analysed during the current study are available in the Figshare repository https://doi.org/10.6084/m9.figshare.12044670[56]. The raw data that support the findings of this study are available from the corresponding author upon request. These data may not be posted publicly due to privacy restrictions. For interested readers, an alternative open data API service with global EV charging infrastructure data is available from OpenChargeMap (https://openchargemap.org/), which is derived from a variety of public sources and contributions. Source Data are provided with this paper.

Code availability

All custom code and algorithm replication materials have been deposited on the Github repository using Zenodo version releases at https://doi.org/10.5281/zenodo.1419830.

Received: 14 February 2019; Accepted: 2 April 2020;
Published online: 01 June 2020

References

1. *Market Data: EV Market Forecasts: Global Forecasts for Light Duty Plug-In Hybrid and Battery EV Sales and Populations: 2018–2030* (Navigant, 2019).
2. *Inventory of US Greenhouse Gas Emissions and Sinks: 1990-2016* Document No. 430-R-18-003 (EPA, 2018).
3. The electric battery vehicle. *Nature* **144**, 627 (1939).
4. Michalek, J. J. et al. Valuation of plug-in vehicle life-cycle air emissions and oil displacement benefits. *Proc. Natl Acad. Sci. USA* **108**, 16554–16558 (2011).
5. Tessum, C. W., Hill, J. D. & Marshall, J. D. Life cycle air quality impacts of conventional and alternative light-duty transportation in the United States. *Proc. Natl Acad. Sci. USA* **111**, 18490–18495 (2014).
6. Holland, S. P., Mansur, E. T., Muller, N. Z. & Yates, A. J. Are there environmental benefits from driving electric vehicles? The importance of local factors. *Am. Econ. Rev.* **106**, 3700–3729 (2016).
7. Li, S., Tong, L., Xing, J. & Zhou, Y. The market for electric vehicles: indirect network effects and policy design. *J. Assoc. Environ. Resour. Econ.* **4**, 89–133 (2017).
8. Andreoni, J. Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ. J.* **100**, 464–477 (1990).
9. Kotchen, M. Green markets and private provision of public goods. *J. Political Econ.* **114**, 816–834 (2006).
10. Warner, M. & Amir, H. Managing markets for public service: the role of mixed public–private delivery of city services. *Public Adm. Rev.* **68**, 155–166 (2008).
11. Warner, M. & Hebdon, R. Local government restructuring: privatization and its alternatives. *J. Policy Anal. Manag.* **20**, 315–336 (2001).
12. Carley, S., Krause, R. M., Lane, B. W. & Graham, J. D. Intent to purchase a plug-in electric vehicle: a survey of early impressions in large US cites. *Transp. Res. Part D* **18**, 39–45 (2013).
13. Sovacool, B. K. & Hirsh, R. F. Beyond batteries: an examination of the benefits and barriers to plug-in hybrid electric vehicles (PHEVS) and a vehicle-to-grid (V2G) transition. *Energy Policy* **37**, 1095–1103 (2009).
14. Helveston, J. P. et al. Choice at the pump: measuring preferences for lower-carbon combustion fuels. *Environ. Res. Lett.* **14**, 084035 (2019).
15. Xu, Y., Çolak, S., Kara, E. C., Moura, S. J. & González, M. C. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nat. Energy* **3**, 484–493 (2018).
16. Asensio, O. I. Correcting consumer misperception. *Nat. Energy* **4**, 823–824 (2019).
17. Williams, B. & DeShazo, J. Pricing workplace charging: financial viability and fueling costs. *Transp. Res. Rec.* **2454**, 68–75 (2014).
18. Asensio, O. I. & Delmas, M. A. Nonprice incentives and energy conservation. *Proc. Natl Acad. Sci. USA* **112**, E510–E515 (2015).
19. Asensio, O. I. & Delmas, M. A. The dynamics of behavior change: evidence from energy conservation. *J. Econ. Behav. Organ.* **126**, 196–212 (2016).

20. Alexander, L., Jiang, S., Murga, M. & González, M. C. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C* **58**, 240–250 (2015).
21. González, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
22. Kim, Y. Convolutional neural networks for sentence classification. In *Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (eds Moschitti, A. et al.) 1746–1751 (Association for Computational Linguistics, 2014).
23. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
24. Socher, R. et al. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proc. 2013 Conference on Empirical Methods in Natural Language Processing* (eds Yarowsky, D. et al.) 1631–1642 (2013).
25. dos Santos, C. & Gatti, M. Deep convolutional neural networks for sentiment analysis of short texts. In *Proc. COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (eds Tsujii, J. & Hajic, J.) 69–78 (Dublin City University and Association for Computational Linguistics, 2014).
26. Yih, W.-t., Toutanova, K., Platt, J. C. & Meek, C. Learning discriminative projections for text similarity measures. In *Proc. Fifteenth Conference on Computational Natural Language Learning* (ed. Pradhan, S.) 247–256 (Association for Computational Linguistics, 2011).
27. Shen, Y., He, X., Gao, J., Deng, L. & Mesnil, G. Learning semantic representations using convolutional neural networks for web search. In *Proc. 23rd International Conference on World Wide Web* (ed. Chung, C.-W.) 373–374 (ACM, 2014).
28. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. A convolutional neural network for modelling sentences. In *Proc. 52nd Annual Meeting of the Association for Computational Linguistics* (eds Toutanova, K. and Wu, H.) 655–665 (Association for Computational Linguistics, 2014).
29. Collobert, R. et al. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **12**, 2493–2537 (2011).
30. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. in *Advances in Neural Information Processing Systems* (eds Burges, C. J. C. et al.) 3111–3119 (Curran Associates, Inc., 2013).
31. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
32. Yin, W., Kann, K., Yu, M. & Schütze, H. Comparative study of CNN and RNN for natural language processing. Preprint at http://arXiv.org/abs/arXiv:1702.01923 (2017).
33. Li, J., Chen, X., Hovy, E. & Jurafsky, D. Visualizing and understanding neural models in NLP. In *Proc. 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (eds Knight, K. et al.) 681–691 (Association for Computational Linguistics, 2016).
34. Tixier, A. J.-P. Notes on deep learning for NLP. Preprint at http://arXiv.org/abs/arXiv:1808.09772 (2018).
35. Song, B., Lee, C., Yoon, B. & Park, Y. Diagnosing service quality using customer reviews: an index approach based on sentiment and gap analyses. *Serv. Bus.* **10**, 775–798 (2016).
36. Wang, H., Can, D., Kazemzadeh, A., Bar, F. & Narayanan, S. A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle. In *Proc. ACL 2012 System Demonstrations* (ed. Zhang, M.) 115–120 (Association for Computational Linguistics, 2012).
37. Hopkins, D. J. & King, G. A method of automated nonparametric content analysis for social science. *Am. J. Political Sci.* **54**, 229–247 (2010).
38. Bail, C. A. The cultural environment: measuring culture with big data. *Theory Soc.* **43**, 465–482 (2014).
39. Ulibarri, N., Scott, T. A. & Perez-Figueroa, O. How does stakeholder involvement affect environmental impact assessment? *Environ. Impact Assess. Rev.* **79**, 106309 (2019).
40. Board, T. R. & Council, N. R. *Overcoming Barriers to Deployment of Plug-in Electric Vehicles* (National Academies Press, 2015).
41. Gerardo Zarazua de Rubens, L. N. & Sovacool, B. K. Dismissive and deceptive car dealerships create barriers to electric vehicle adoption at the point of sale. *Nat. Energy* **3**, 501–507 (2018).
42. Rezvani, Z., Jansson, J. & Bodin, J. Advances in consumer electric vehicle adoption research: a review and research agenda. *Transp. Res. Part D* **34**, 122–136 (2015).
43. *Plug-In Electric Vehicle Deployment Policy Tools: Zoning, Codes, and Parking Ordinances* Technology Bulletin (DOE, 2015).
44. *National Plug-In Electric Vehicle Infrastructure Analysis* (DOE, 2017).
45. DeShazo, J. R. Improving incentives for clean vehicle purchases in the United States: challenges and opportunities. *Rev. Environ. Econ. Policy* **10**, 149–165 (2016).
46. DeShazo, J., Sheldon, T. L. & Carson, R. T. Designing policy incentives for cleaner technologies: lessons from California's plug-in electric vehicle rebate program. *J. Environ. Econ. Manag.* **84**, 18–43 (2017).
47. Zhang, Y. & Wallace, B. C. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In *Proc. Eighth International Joint Conference on Natural Language Processing* (eds Kondrak, G. & Watanabe, T.) 253–263 (Asian Federation of Natural Language Processing, 2017).
48. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
49. Karpathy, A., Johnson, J. & Fei-Fei, L. Visualizing and understanding recurrent networks. In *International Conference on Learning Representations 2016 Workshop* (International Conference on Learning Representations, 2016).
50. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
51. Papke, L. E. & Wooldridge, J. M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J. Appl. Econ.* **11**, 619–632 (1996).
52. Papke, L. E. & Wooldridge, J. M. Panel data methods for fractional response variables with an application to test pass rates. *J. Econ.* **145**, 121–133 (2008).
53. Ramalho, E. A., Ramalho, J. J. & Murteira, J. M. Alternative estimating and testing empirical strategies for fractional regression models. *J. Econ. Surv.* **25**, 19–68 (2011).
54. *2010 Census Urban and Rural Classification and Urban Area Criteria* (US Census, 2010).
55. Conneau, A., Schwenk, H., Barrault, L. & Lecun, Y. Very deep convolutional networks for text classification. In *Proc. 15th Conference of the European Chapter of the Association for Computational Linguistics* Vol. 1 (eds Lapata, M. et al.) 1107–1116 (Association for Computational Linguistics, 2017).
56. Asensio, O. I. & Ha, S. *Trained CNN and LSTM model on EV charging station reviews for sentiment analysis* (Figshare, 2020); https://doi.org/10.6084/m9.figshare.12044670

Acknowledgements

This research was supported by a grant from the National Science Foundation (CPS award no. 1931980), the Civic Data Science REU programme at Georgia Tech (NSF award no. IIS-1659757), the Anthony and Jeanne Pritzker Family Foundation, the Sustainable LA Grand Challenge. We are grateful to E. Zegura and C. Le Dantec for discussions. For valuable research assistance, we thank M. E. Burke, S. Dharur, S. Oh and D. Marchetto. Special thanks to N. Hajjar.

This research was also supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA.

Author contributions

O.I.A. directed the research and wrote the paper; A.D., E.W. and K.A. developed code, analysed data and wrote the paper; K.A., C.H. and S.H. implemented algorithms and performed experiments; S.H. investigated model validation and interpretability. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

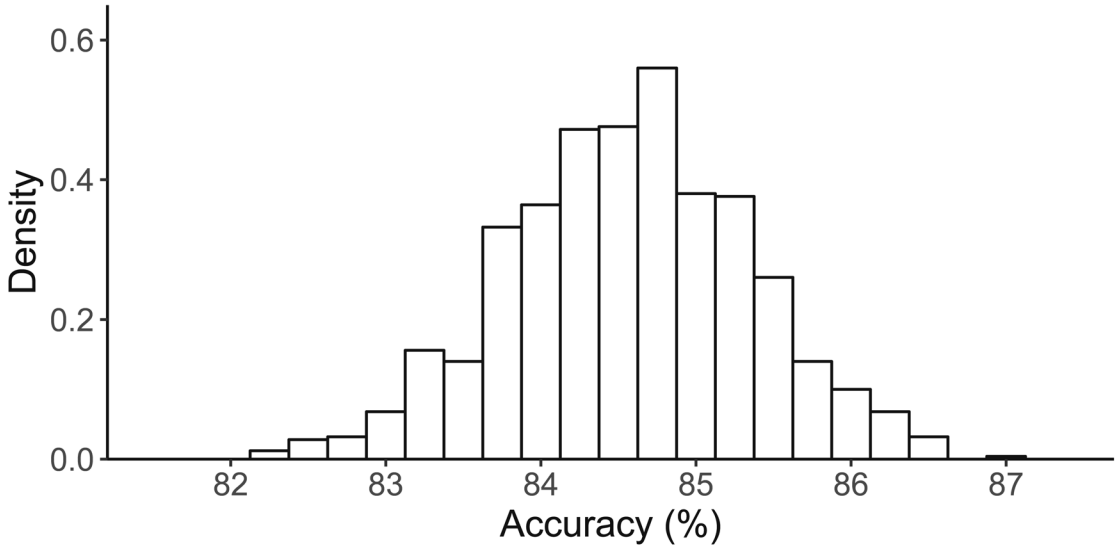
Extended data is available for this paper at https://doi.org/10.1038/s41893-020-0533-6.

Supplementary information is available for this paper at https://doi.org/10.1038/s41893-020-0533-6.

Correspondence and requests for materials should be addressed to O.I.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Extended Data Fig. 1 | Distribution of CNN classifier predictions for 1,000 model runs. The mean test accuracy for 1,000 runs is 84.6% with a S.D. of 0.79.

	Public	Private	Total
Positive	11,761	54,094	65,855
Negative	10,852	50,550	61,402
Total	22,613	104,644	127,257

Extended Data Table 1 | Descriptive statistics, public and private. Counts of machine classified reviews of binary sentiment by public and private ownership. 2,256 reviews were submitted in locations where it was impossible to discern whether it was public or private.

	Review Rate			Negativity Score		
	FRM	*FRM*	*FRM*	*FRM*	*FRM*	*FRM*
	(I)	(II)	(III)	(IV)	(V)	(VI)
Geographical Area						
Urban	-0.021**	-0.035***	-0.035***	0.147***	0.125***	0.125***
	(0.008)	(0.007)	(0.007)	(0.013)	(0.012)	(0.012)
Non-Urban	0.025**	0.017*	0.017*	0.018	0.007	0.007
	(0.012)	(0.010)	(0.010)	(0.016)	(0.015)	(0.015)
Type of Location						
Public	-0.010	-0.006		0.004	0.011	
	(0.013)	(0.015)		(0.015)	(0.015)	
Government			-0.011			0.011
			(0.016)			(0.016)
Station Characteristics						
Number of Connectors	-0.082***	-0.077***	-0.077***	-0.013***	-0.008*	-0.008*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Number of Networks	-0.011	-0.010	-0.010	0.026*	0.016	0.016
	(0.014)	(0.018)	(0.017)	(0.016)	(0.018)	(0.018)
Quality Rating		-0.036***	-0.036***		-0.049***	-0.049***
		(0.002)	(0.002)		(0.002)	(0.002)
Points of Interest						
Residential	0.063*	0.032	0.030	0.072 ***	0.032	0.032
	(0.035)	(0.041)	(0.041)	(0.025)	(0.021)	(0.021)
Shopping	-0.107***	-0.099***	-0.101***	0.041 ***	0.050***	0.050***
	(0.011)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Restaurants	-0.017	-0.009	-0.011	-0.010	0.001	-0.001
	(0.014)	(0.015)	(0.015)	(0.017)	(0.017)	(0.017)
Healthcare	0.040**	0.022	0.020	0.024	0.005	0.005
	(0.017)	(0.025)	(0.025)	(0.019)	(0.022)	(0.023)
Hotel and Lodging	0.058***	0.054***	0.052 ***	-0.069***	-0.073***	-0.073***
	(0.012)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Workplace	0.027*	0.025	0.022	0.001	-0.002	-0.002
	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)	(0.016)
Supermarket	-0.076***	-0.070***	-0.072***	0.078 ***	0.087***	0.087***
	(0.012)	(0.015)	(0.015)	(0.017)	(0.016)	(0.017)
Car Dealership	0.031***	0.023*	0.020	0.042 ***	0.035**	0.035**
	(0.010)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Education	0.055***	0.042***	0.036**	0.024	0.006	0.014
	(0.015)	(0.016)	(0.018)	(0.019)	(0.017)	(0.018)
Entertainment	0.007	0.010	0.008	0.022	0.03	0.025
	(0.017)	(0.018)	(0.019)	(0.020)	(0.020)	(0.020)
Convenience and Gas Station	-0.001	-0.002	-0.004	0.006	0.005	0.005
	(0.011)	(0.014)	(0.015)	(0.016)	(0.017)	(0.017)
Transit Station	-0.041***	-0.034**	-0.042 **	0.017	0.023	0.034*
	(0.016)	(0.016)	(0.017)	(0.018)	(0.016)	(0.018)
RV Park	0.186***	0.155***	0.153***	-0.080 ***	-0.129***	-0.129***
	(0.021)	(0.019)	(0.019)	(0.031)	(0.036)	(0.036)
Outdoor	0.007	0.010	0.002	-0.034 *	-0.030 *	-0.018
	(0.027)	(0.025)	(0.026)	(0.020)	(0.018)	(0.019)
Airport	0.017	0.015	0.007	0.008	0.007	0.019
	(0.018)	(0.018)	(0.019)	(0.041)	(0.038)	(0.039)
Services	0.005	0.012	0.010	-0.035	-0.024	-0.024
	(0.022)	(0.022)	(0.022)	(0.025)	(0.023)	(0.023)
Place of Worship	-0.051	-0.042	-0.044	0.004	0.016	0.016
	(0.081)	(0.074)	(0.074)	(0.025)	(0.025)	(0.025)
Shopping Center	0.012	0.031	0.029	-0.080 ***	-0.037*	-0.037*
	(0.085)	(0.068)	(0.068)	(0.021)	(0.022)	(0.022)
Library	0.021	0.021	0.013	0.039	0.037	0.049*
	(0.019)	(0.022)	(0.023)	(0.027)	(0.024)	(0.026)
Street Parking	-0.008	-0.014	-0.022	0.077 **	0.062**	0.074**
	(0.025)	(0.023)	(0.024)	(0.032)	(0.032)	(0.033)
Visitor Center	-0.044	-0.025	-0.028	-0.011	0.009	0.009
	(0.028)	(0.027)	(0.027)	(0.024)	(0.024)	(0.025)
Car Rental	0.235***	0.192***	0.190***	0.278 ***	0.217**	0.217**
	(0.042)	(0.037)	(0.038)	(0.085)	(0.088)	(0.088)
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	0.206	0.206	0.048	0.105	0.120
Note:						*p<0.1; **p<0.05; ***p<0.01

Extended Data Table 2 | Main results. FRM results for the review rate and negativity score.

	Rural	Urban Cluster	Urban Center	Total
Positive	4,745	4,819	56,291	65,855
Negative	2,509	2,419	56,402	61,402
Total	7,254	7,310	112,693	127,257

Extended Data Table 3 | Descriptive statistics, urban and rural. Counts of machine classified reviews of binary sentiment by geographic area type as defined by U.S. Census designations.

State	Public			Private			No. of Reviews
	Free	Paid	p-value	Free	Paid	p-value	
Los Angeles-Long Beach-Anaheim, CA	0.52	0.55	0.36	0.59	0.55	0.11	22,878
San Francisco-Oakland-Hayward, CA	0.45	0.47	0.73	0.59	0.54	0.11	8,951
Atlanta-Sandy Springs-Roswell, GA	0.48	0.59	0.23	0.49	0.53	0.32	5,442
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.50	0.46	0.66	0.47	0.46	0.76	3,452
Phoenix-Mesa-Scottsdale, AZ	0.14	0.50	0.21	0.47	0.55	0.06	2,863
New York-Newark-Jersey City, NY-NJ-PA	0.50	0.49	0.90	0.51	0.45	0.16	2,060
Chicago-Naperville-Elgin, IL-IN-WI	0.46	0.60	0.20	0.51	0.47	0.32	1,781
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.44	0.43	0.96	0.49	0.57	0.18	1,586
Boston-Cambridge-Newton, MA-NH	0.44	0.58	0.12	0.43	0.50	0.18	1,438
Dallas-Fort Worth-Arlington, TX	0.48	0.61	0.38	0.49	0.60	0.05	1,139
Nashville-Davidson-Murfreesboro-Franklin, TN	0.63	0.60	0.88	0.41	0.44	0.72	1,082
Denver-Aurora-Lakewood, CO	0.64	0.63	0.95	0.43	0.40	0.60	1,042
Detroit-Warren-Dearborn, MI	0.49	0.49	1.00	0.46	0.44	0.76	792
Minneapolis-St. Paul-Bloomington, MN-WI	0.33	0.63	0.01	0.43	0.35	0.27	658
Austin-Round Rock, TX	0.55	0.25	0.03	0.37	0.39	0.84	508
Hartford-West Hartford-East Hartford, CT	0.38	0.50	0.62	0.45	0.49	0.76	471
Kansas City, MO-KS	0.28	0.49	0.58	0.39	0.32	0.57	488
Chattanooga, TN-GA	0.29	0.42	0.67	0.50	0.68	0.26	292

Extended Data Table 4 | Probability of negative sentiment for 18 core-based statistical areas in the United States. Results of t-tests for free and paid stations by public and private ownership in 18 CBSAs in the United States.

State	Public			Private			No. of Reviews
	Free	Paid	p-value	Free	Paid	p-value	
California	0.46	0.49	0.16	0.50	0.52	0.13	54,684
Washington	0.50	0.46	0.45	0.40	0.46	0.06	7,830
Oregon	0.47	0.53	0.56	0.36	0.43	0.04	7,027
Georgia	0.49	0.49	0.98	0.42	0.52	0.00	6,623
Florida	0.38	0.43	0.36	0.48	0.47	0.75	4,420
Maryland	0.41	0.43	0.80	0.44	0.41	0.53	3,541
Arizona	0.45	0.51	0.77	0.40	0.52	0.00	3,365
New York	0.45	0.49	0.56	0.41	0.47	0.09	2,894
Texas	0.54	0.47	0.33	0.46	0.51	0.12	2,615
Virginia	0.48	0.42	0.59	0.41	0.48	0.08	2,588
Pennsylvania	0.44	0.57	0.28	0.43	0.49	0.24	2,536
North Carolina	0.42	0.46	0.65	0.41	0.52	0.07	2,181
Colorado	0.40	0.53	0.25	0.39	0.37	0.73	2,161
Illinois	0.54	0.55	0.92	0.49	0.46	0.45	2,105
Massachusetts	0.42	0.52	0.23	0.41	0.46	0.30	2,074
Tennessee	0.53	0.57	0.69	0.43	0.51	0.10	1,983
Michigan	0.34	0.39	0.61	0.39	0.29	0.08	1,488
Ohio	0.35	0.50	0.29	0.43	0.51	0.17	1,443
New Jersey	0.48	0.49	0.94	0.48	0.50	0.78	1,390
Hawaii	0.53	0.71	0.45	0.57	0.57	0.93	1,259

Extended Data Table 5 | Probability of negative sentiment for top 20 U.S. states. Results of t-tests for free and paid stations by public and private ownership in the top 20 states by number of reviews.

Human Labeled Sentiment	Count	Percentage (%)
Positive	4,933	55.1
Negative	4,020	44.9
Total	8,953	100.0

Extended Data Table 6 | Balance of training data. Counts of positive and negative reviews by two human annotators (κ=0.84).