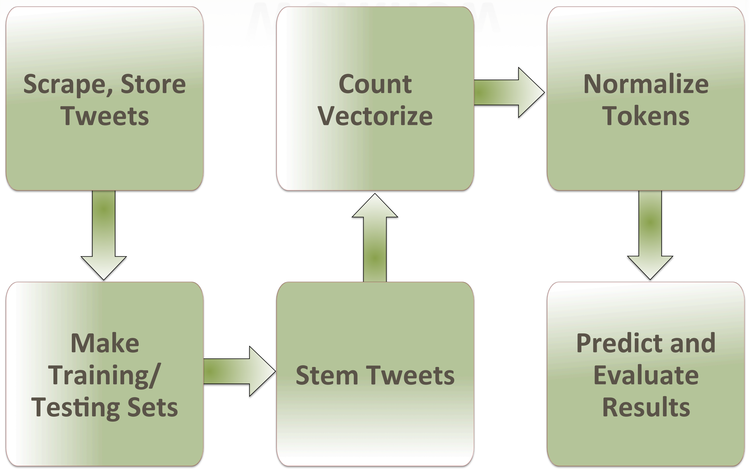
Scrapable Insights: How Twitter Can Inform on Kenyan Electrical Grid Growth and Development

†Shayne T. O'Brien, ‡Margaret Sy, ‡Dr. Eric Brewer
†State University of New York at Geneseo;
‡University of California, Berkeley

Abstract

While reports have shown that Kenyan electrical connectivity rates have increased significantly over the past few years, there is little known about how the electric grid and its reliability have changed over time. In this study, the research team utilizes a Twitter API to scrape tweets from the national electric company Kenya Power and Lighting Company (KPLC). Using statistical machine learning algorithms, these tweets are given a binary classification as to whether or not they are power outage related. Outage related results are visualized and analyzed in an effort to determine where outages are occurring as well as gain insight into what can be learned via the public customer service communication medium.

Workflow



Acknowledgements

1. Funded by NSF Grant No. 1359499.
2. UC Berkeley SUPERB-ITS REU.
3. Dr. Eric Brewer, PI.

Results and Discussion

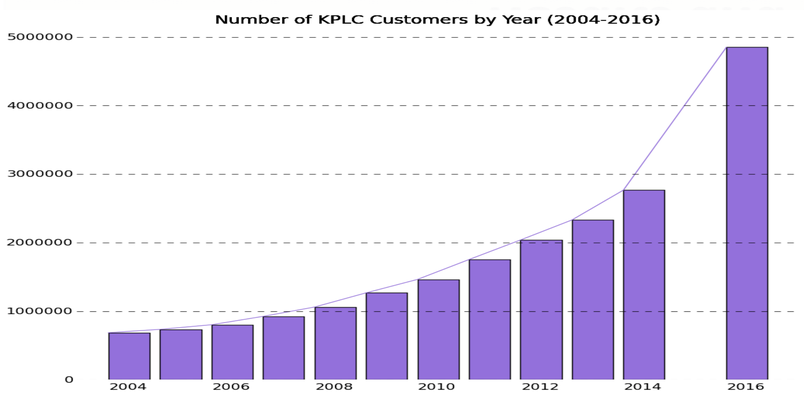


Figure 1. Number of Kenya Power and Lighting customers between 2004 and 2016.

According to Kenya Power and Lighting Company (KPLC), the company's number of customers has increased drastically since 2004. At the same time, very little data has been released about the reliability and growth of the grid over time. By scraping publically available tweets from KPLC's Twitter account and classifying them as outage-related or not using machine learning, we can gain valuable insight into this area of study. Preliminary results show a strong bias toward densely populated areas and time of day. This implies that normalization of results will be necessary in order to understand the big picture. Future work includes implementing a neural network for classifying tweets, improving location parsing, and adding functionality to interpret tweets in Swahili. End results will be able to inform on how Kenya's electrical grid has changed between 2013, when data collection began, and present day. These results may inspire future studies or improvements to the grid.

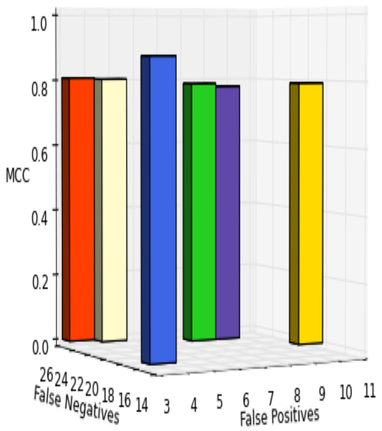


Figure 2. Support vector machine algorithm with linear kernel is the best machine learning classifier of the tested algorithms (several not shown).

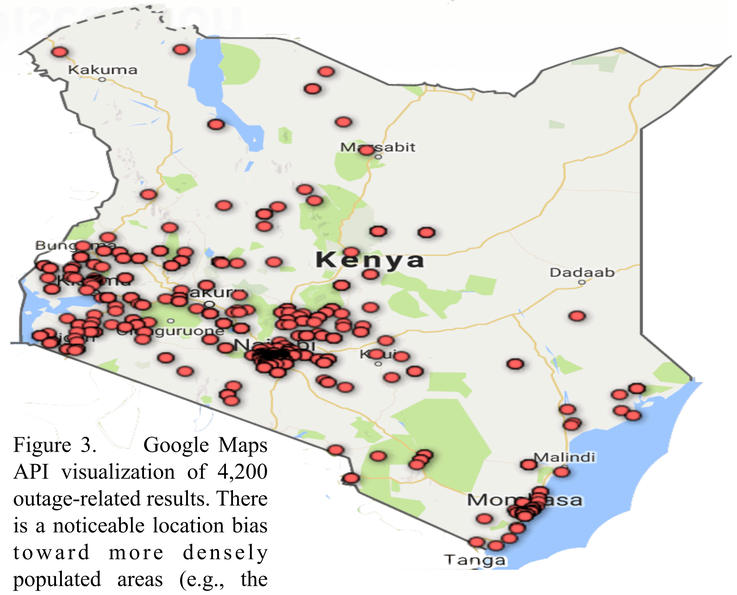


Figure 3. Google Maps API visualization of 4,200 outage-related results. There is a noticeable location bias toward more densely populated areas (e.g., the capital city of Nairobi).

Error Type	Actual	Predicted	Tweet Text
FN	['L2']	[ ]	@kenyapower thank you very much lights are back kudos to your team..
FP	[ ]	['L1']	@KenyaPower is there a reason you schedule black outs everyday in this area
FN	['L1']	[ ]	@KenyaPower Hi same story as yesterday and it's now total blackout.. Please attend

Figure 4. Examples of misclassifications by the support vector machine algorithm using a linear kernel.

Conclusion

This research project is an example of how publically available information can be used, by anyone, to inform on issues relating to global connectivity. By scraping tweets using a Twitter API, applying machine learning algorithms to classify the data, and analyzing the results, the research team is attempting to fill in the gaps in knowledge about how the electrification of Kenya is progressing.