Data Science for Social Media in Emergency Management

New Mexico
Supercomputing Challenge
Final Report
April 4, 2018

Team PESSF118
Pinon Elementary School, Santa Fe

Team Members “hacker names” (real names withheld for privacy):
SoccerChamp
volleyballqueen
SkaterAssistancedDog
SpursQueen
DarkDJ
War machine

Teacher:
Delara Sharma

Mentor:
Joel Berendzen (GenerisBio, LLC)

Executive Summary

Our project is about using machine learning to classify Twitter streams for use in emergency management situations. We chose this problem because experts on the use of Social Media in Emergency Management (SMEM) helped us see how important the problem is, and few others seem to be working on it. We chose Twitter because it is the social media site that SMEM workers find most useful.

The problem we chose was to collect data on pet ownership on a set of Twitter accounts and train a classifier (pet owner/not pet owner) from this data set. Pet ownership was chosen for three reasons: (1) we like pets, (2) pet ownership is not something that Twitter account owners generally feel privacy concerns about or would lie about, and (3) it's not easy to tell from an individual Tweet whether the account owner owns a pet. This problem seems hard enough to stand in for more difficult problems such as whether or not the reports from a Twitter account can be trusted in an emergency.

We were not able to complete the project, but we intend to continue it next year.

Our supercomputing approach is to use data supercomputing on Amazon Web services. We are using serverless python (Amazon Lambda) so that our computations may be scaled cheaply on demand without us having to work hard at web site deployment.

Our first accomplishment was creating our own website, https://eaglebytes.org. We then built a machine learning application (funyun) for public use. **Our greatest accomplishment was giving an oral presentation** on our project to outside reviewers. Our next step will be to collect data that can be used for training our machine learning classifiers.

We learned a lot about how a team works to create software, and also a bit about machine learning. One problem area for us is that we cannot use many public web sites due to age restrictions in the terms of service, while the computers we have available (Chromebooks) and are locked down tight for classroom use. We found workarounds for this year using a shared development server (Eclipse Che). For next year we were lucky enough to obtain a grant that gets us Chromebooks with local storage for development use.

Statement of the problem

We were moved to work on social media in disasters after watching a lot of disasters unfold on TV over the summer. We were especially moved by watching people rescue pets, because we love pets.

Rescue team rescues dogs from disaster area.

After talking with an SMEM worker, Marlita Reddy-Hjelmfelt, we learned three problems that people using social media in emergencies face:

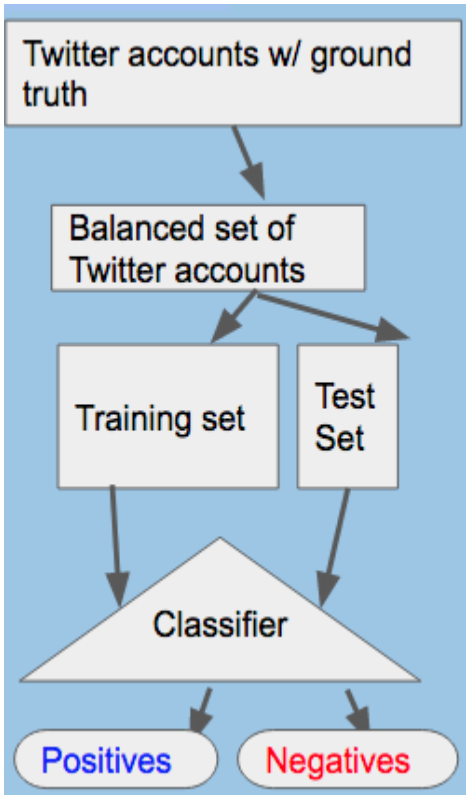
1. **Finding good search terms to retrieve relevant posts.** Many posts do not include standard hashtags. Few posts are geolocated.
2. **Doing complex searches on streams**. VOST workers rely on a combination of mental filtering and personal requests to friendly academics".
3. **Distinguishing false from true reports.** An emergency attracts some people who falsify reports, which have to be distinguished from real reports using human judgement.

We found scholarly articles on SMEM [1-5], and found these observations seem consistent with what has been published on SMEM.

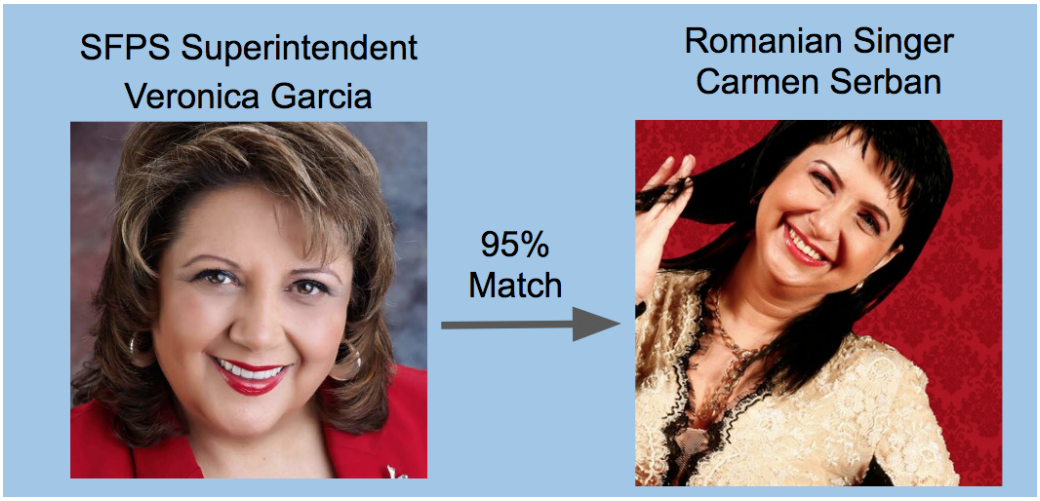
Our mentor pointed out the **machine learning** is a hot area for computational science now and that 5th-graders can actually do many things with machine learning by following examples. **We chose as our hypothesis that machine learning is useful for classifying SMEM posts.** The specific problem we chose to test this hypothesis is to see how well we can classify Twitter accounts of pet-owners from non-pet-owners.

Methods we used to solve the problem

The approach we used is to clone, hack and copy code so the work is easier. The tasks we have to do to implement machine learning on Twitter accounts is shown below:

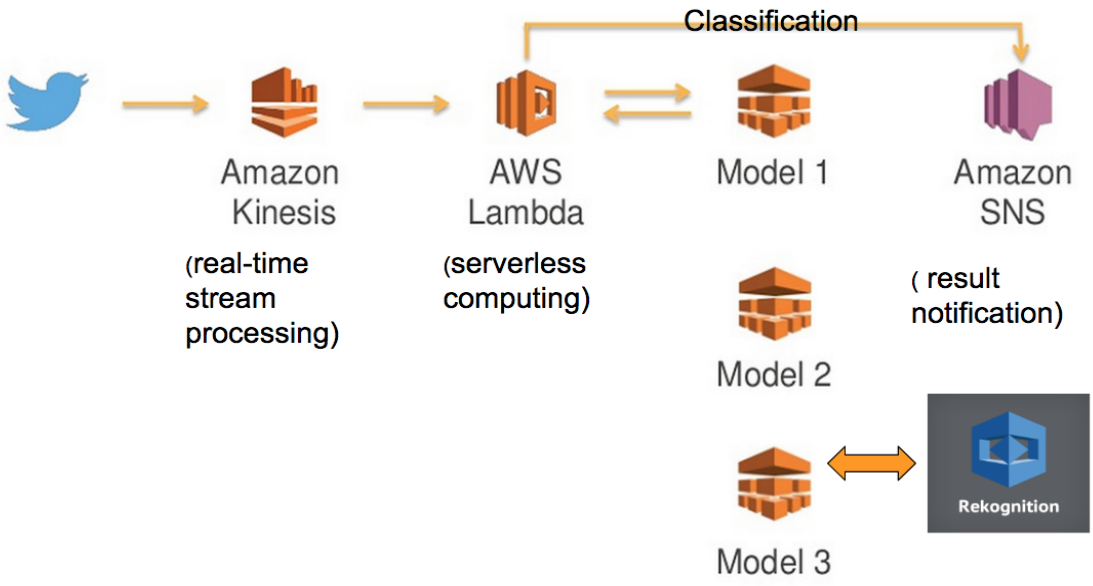


The first step is to collect Twitter accounts with ground truth attached. We found in the course of testing the already-trained classifiers in Amazon Rekognize that it is entertaining to look at the output of the classifiers, particularly the one that assesses whether or not a picture looks like a celebrity. We hope to use this entertainment value to convince people to give us their Twitter account handles together with whether or not they own pets. An example using our Superintendent is shown below:



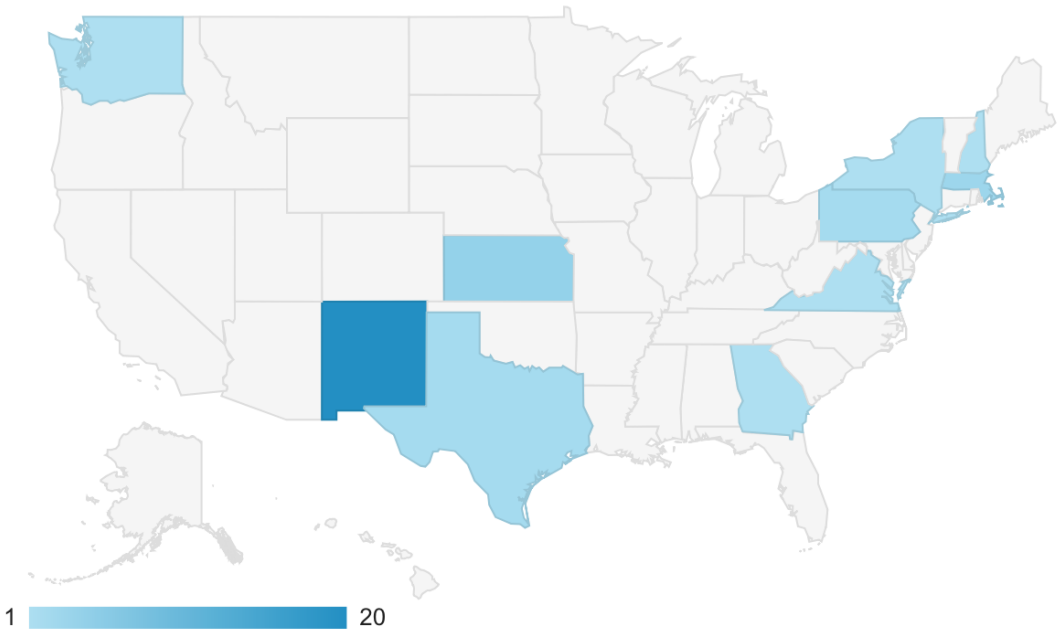
How we verified and validated our model

We did not reach the point of building our model, much less verifying and validating it. The figure below shows the data-flow architecture that our models live in. We plan to have model 1 be a simple word-frequency model on individual Tweets, while model 2 will be a machine-learning model that operates on the entire Twitter stream. Model 3 will have image analysis (through Amazon Rekognition) added and employ deep learning.



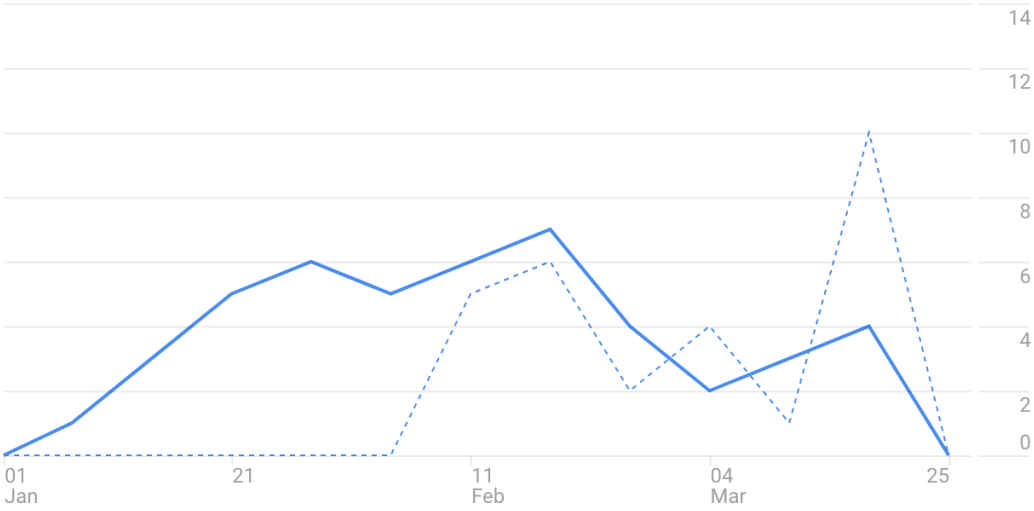
Results of study

We have learned some standard software development techniques, such as working with source code repositories and issuing pull requests. Together we have only written a little over 200 lines of code. Our code gets an A for style and about 70% for test coverage. We are using GitHub to keep track of code. We have made progress on our website. It now uses trained machine-learning classifiers in Amazon Rekognize to recognize food, pets, and feature of faces (e.g, gender, emotional state, age, and whether the person wears glasses) We have also had people from other states go onto our website as shown below from our Google Analytics data (some we think to hack it).



This is a map of where we have other users or hackers (Google Analytics).

Users	Sessions	Bounce Rate	Session Duration
28	60	40%	8m 7s
↑16.7%	↑50%	↓44.8%	↑108%



This is a plot of web site visitors over time (Google Analytics).

Conclusions we reached in analyzing our results

We have used good software practices in implementing *funyun,* including they have used appropriate software engineering tools (pypi, Travis, codecov, pyup, and codacy). We have had over 28 users in are github website. We have also had 60 sessions.

Software, references, tables, and other products of our work--Logan

Our code lives in a source code repository at https://github.com/EagleBytes2017/funyun.

References:

1. Petronzio, Matt (2017) "Facebook opens its Community Help API to disaster organizations". *Mashable*, November 29, 2017. *http://mashable.com/2017/11/29/facebook-community-help-api-fundraising*
2. Zagorecki, Adam & Johnson, David & Ristvej, Jozef. (2013) "Data mining and machine learning in the context of disaster and crisis management". *Int. J. Emerg. Mgmt.* **9**.351-365. DOI:10.1504/IJEM.2013.059879.
3. Ilyas, Andrew. (2014) "MicroFilters: Harnessing twitter for disaster management". Published in *Global Humanitarian Technology Conference*, 2014 IEEE. DOI:10.1109/GHTC.2014.6970316
4. Bruns, Axel & Burgess, Jean (2014), "Crisis Communication in Natural Disasters: The Queensland Flood and Christchurch Earthquakes". In Weller, Katrin, Bruns, Axel, Burgess, Jean, Mahrt, Merja, & Puschmann, Cornelius (Eds.) *Twitter and Society* Peter Lang, New York, pp. 373-384. http://eprints.qut.edu.au/66329
5. Smith, Brian (2010), "Socially distributing public relations: Twitter, Haiti, and interactivity in social media". *Public Relations Rev.* **36**:329-335. DOI: 10.1016/j.pubrev.2010.08.005

Our most significant achievement on the project

We believe our most significant achievement was to learn to give oral presentations and answer questions from adults about what we did.

Acknowledgment of the people and organizations that helped us

We wish to acknowledge **Marlita Reddy-Hjelmfeldt** for helping us understand the problems real SMEM workers face when they work on an emerency. **M. Jay Mitchell**, New Mexico Cabinet Secretary for Homeland Security and Emergency Management showed early enthusiasm and put us in contact with **Karen Takai**, Public Information Officer for NMDHSEM, who put us in contact with Ms. Redd-Hjelmfeldt. **Ms. Felicia Maestas**, our Digital Learning Coordinator at Pinon Elementary, was always ready to step in and help us. Most of all, we want to thank **our parents** for their continuing support through the duration of our project.