

CS-510 Project Presentation

Adit Gupta & Coire Gavin-Hanner

Our aim is to develop an **AI system** that can **extract usable information** from the natural language in a review or **social media post**. We will be focusing on a small subset of that problem. Given a review or post resulting from a search for an experience the AI system will be asked to determine if the post indeed concerns the experience in question. The experience that we will be focusing on is watching the **Eagles game at a bar**. Our goal will be to use the AI system to check if a review was written about an Eagles game or if the review simply contained a word or phrase that matched the search term, such as the word “Eagles”.



The Problem | Impact

Unverified Reviews



Lack of Real Time
Business Analytics



No Efficient Customer
Outreach



5-9%

Increase in Business
Revenue

92%

Use Reviews When
Searching

22%

Revenue loss By 1 Bad
Review



Approach

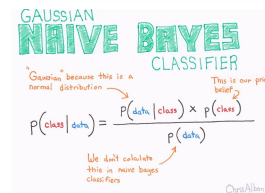
1st

Find Appropriate Data
Source & Normalize
Data



2nd

Create Classifier &
Train with Dataset



3rd

Test Classifier &
analyze results



Data Collection Methods



- Stale data, with long paragraph life structure which is hard to analyze.
- API approval takes a while
- Less metadata available for usage



- Majority of data is ratings, without words (stars)
- Hard to prove is the data is trustworthy
- Expensive to use at scale

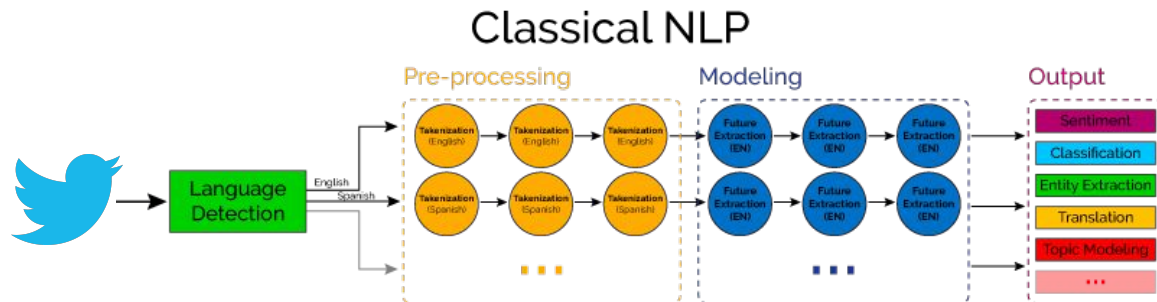


- Short, and structured data via Twitter API
- Geo-Location coordinates
- Less Stale Data

Results

```
((base) Coire@Coires-MacBook-Pro:~/cs510/CS510-Project$python twythonTest.py
```

	user		date	text	favorite_count
2	stephenasmith	Sun Dec 01 21:32:46 +0000	2019	What a disgraceful loss by the @Eagles. You're...	17843
1	ComplexSports	Sun Dec 01 21:34:23 +0000	2019	Eagles fans after losing to the Dolphins https...	12242
0	RealSkipBayless	Sun Dec 01 21:33:57 +0000	2019	HOW BOUT THEM EAGLES?	9717
8	espn	Sun Dec 01 21:33:03 +0000	2019	THE DOLPHINS TAKE DOWN THE EAGLES! 🤔 https://t...	6239
6	RapSheet	Sun Dec 01 21:11:49 +0000	2019	Sometimes, football is weird: The #Bengals are...	6156



Challenges

Challenge 1

- Data Collection
 - Normalize Data
 - Train classifier using dataset
 - Creating file systems to mark specific data items for training

Challenge 2

- Natural Language Understanding
 - For example, understanding whether certain occurrences of words are relevant to the meaning of the keyword, as in “football” and “eagles”

Challenge 3

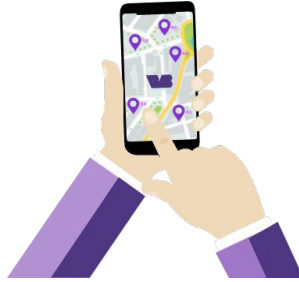
- Real-Time Limitation
 - Our goal was to produce analysis in real time, however, due to the limitation of available data, we were not able to produce real time results.

Future Work

Increase Accuracy



Improve Data Sets



Create User Experience



Team



Coire Gavin-Hanner



Adit Gupta