

Aspect Based Opinion Mining

Gold Team 2

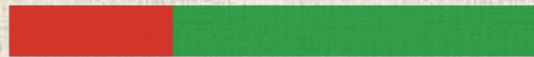
Jithin George Katie Krawczyk Hemanth Natarajan Vineet Tanna Hanjing Wang
georg670@umn.edu krawc034@umn.edu natar060@umn.edu tanna011@umn.edu wang7558@umn.edu
Github: <https://github.com/nhemanthrao23/bigdataproject>

Objective: Leverage Spark MLlib in conjunction with AWS to perform aspect based opinion mining (Topic Modeling with Sentiment Analysis)

“... The wait time was long. However, the hostess was friendly and the waitress came to take our order quickly once we were seated. The chicken tacos were amazing and the sauce on the cheese enchiladas was delicious.”



Food



+0.7



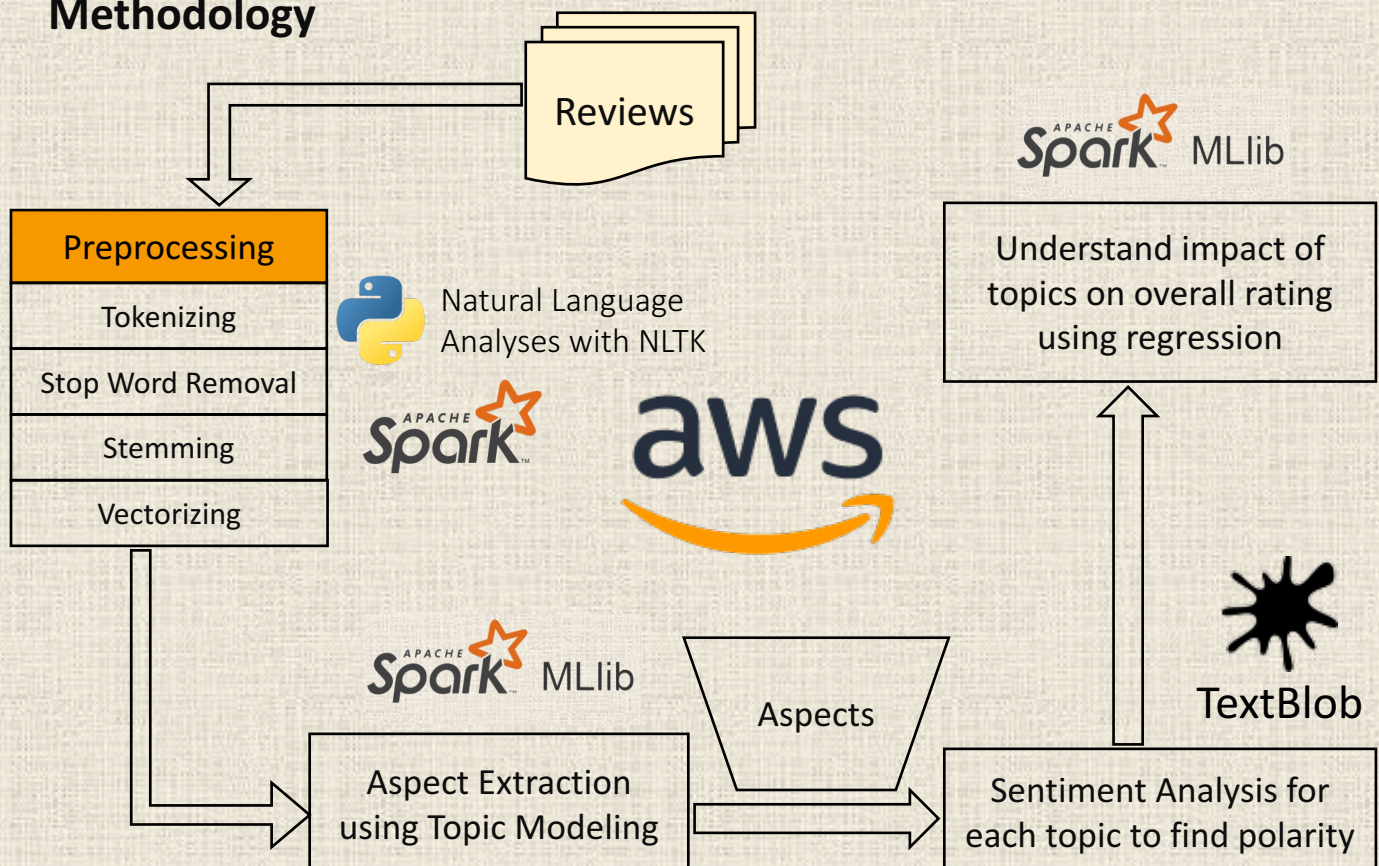
Service



+0.2

Data: 4.7 Million rows of restaurant reviews on Yelp

Methodology



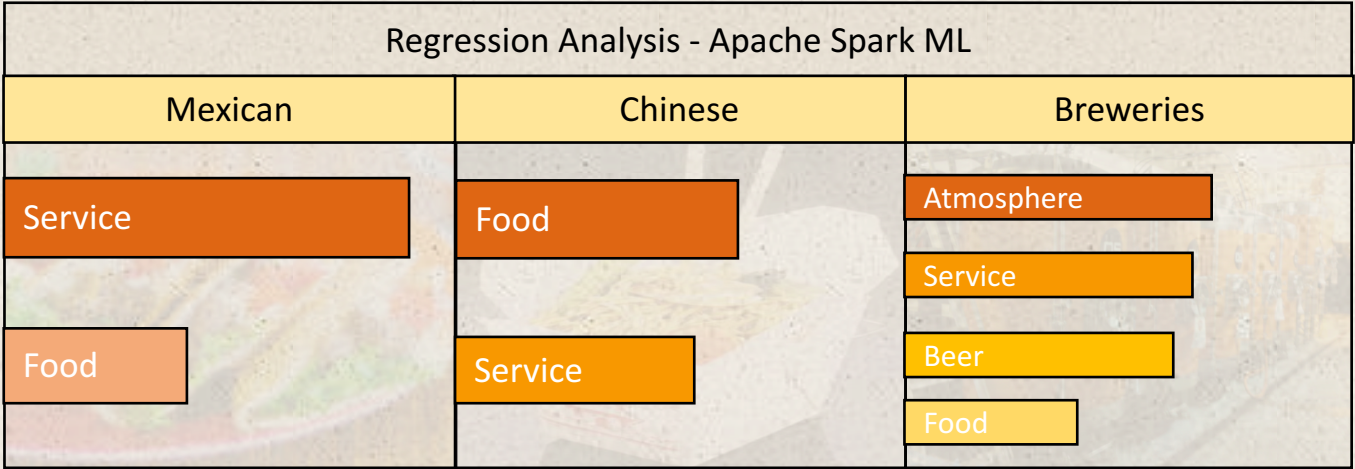
Aspect Based Opinion Mining

Gold Team 2

Results

Topic Modeling - Apache Spark MLlib		
Mexican	Chinese	Breweries
Food - Taco, Sauce, Chicken... Service - Time, Service, Wait...	Food - Noodle, Rice, Chicken... Service - Service, Time, Friendly...	Beer - Beer, Brew, Craft... Service - Time, Wait, Server... Food - Fries, Chicken, Burger... Atmosphere - Friendly, Atmosphere, Staff, Fun...

Sentiment Analysis – TextBlob (based on NLTK)				
	Topic1-Polarity	Topic2-Polarity	..	Topic..N-Polarity
Mexican Village Restaurant	+0.2	+0.5
Taco Shack	+0.2	-0.5
Taqueria El Rey	+0.7	+0.8
Estimate average polarity associated with each topic based on filtered reviews (using aspect keywords) for each restaurant within a cuisine				



Key Takeaways:

- Spark’s ability to integrate Python libraries such as NLTK and TextBlob enables faster computing while performing text mining on large datasets
- AWS’ flexibility in terms of memory expansion and collaboration allows smoother workflow on big data projects
- Spark ML enables users to perform limited machine learning algorithms on large datasets