SOFTWARE REQUIREMENTS:

This project requires Python 2.7 and the following packages:
- NumPy
- Pandas
- Scikit-learn
- Keras
- Xgboost
- Time

FILES:

This project comprises the following files:
- reviews.csv
- capstone_sentiment_analysis.ipynb
- proposal.pdf
- report.pdf
- readme.pdf

DATA:

This project uses the Amazon Fine Food Reviews dataset[1], comprising 568,454 reviews for food products on Amazon up until October 2012. The dataset can be found on reviews.csv. The dataset comprises the following labels:
- Score: an integer ranging from 1 to 5[2], indicating the reviewer's sentiment towards the product;
- Text: text of the review;
- Summary: title of the review;
- Id: integer stating the ordinal position of the review within the dataset
- UserId: unique identifier of the reviewer;
- ProductId: unique identifier of the product reviewed;
- ProfileName: username of the reviewer;
- HelpfulnessNumerator: number of mentions indicating the review was helpful;
- HelpfulnessDenominator: total number of mentions regarding the helpfulness of the review; and
- Time: timestamp of the review.

The following is an example of a review from the dataset:

> Score: 5
> Text: I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.
> Summary: Good Quality Dog Food
> Id: 1
> UserId: A3SGXH7AUHU8GW
> ProductId: B001E4KFG0
> ProfileName: delmartian
> HelpfulnessNumerator: 1
> HelpfulnessDenominator: 1
> Time: 1303862400

AMAZON WEB SERVICES:

Instead of training this model on a local CPU (or GPU), I used Amazon Web Services to launch a P2 EC2 GPU instance.

---

[1] Obtained from Kaggle at https://www.kaggle.com/snap/amazon-fine-food-reviews

[2] A score of 5 being the best.