

Sentiment Classification on Amazon Reviews
Capstone 3 - Project Proposal
Hien Quang

Introduction

The objective of this project is to extract insight from Amazon reviews through the process of sentiment classification

Problem

This project is an exploration of the language and sentiment in Amazon reviews. There are problem could be potentially benefit from this project such as:

- Identify fake reviews
- Identify mistakes when a review is positive but the rating is low or vice versa.
- Identify what is important to customers

Data

Source: <https://nijianmo.github.io/amazon/index.html>

This dataset includes reviews (ratings, text, helpfulness votes), product metadata (descriptions, category information, price, brand, and image features), and links (also viewed/also bought graphs).

Envisioned Approach

Natural Language Processing approach:

- Symbols and punctuation marks removal
- Tokenization
- Stemming
- Lemmatization

Deliverables

- All developed code (in form of Jupyter notebooks)
- Final written report
- Presentation slide deck