Exploratory

PCA

LDA

LSA

Other analysis

Data

Crawl and scrape data with additional info

Look into other sources

Create a command line tool for quickly labeling reviews

Label reviews (agree on some set number)

Supervised Learning

Lit. Review (includes Generative and Discriminative Text Classification with Recurrent Neural Networks)

Design architecture

Train, validate, and test on labeled reviews

Label remaining reviews with optimized model

Complaints Shifts

With aggregate complaints stats from labeled reviews…

Graph complaint shifts year to year

Parallel complaints shift to Turbotax feature introduction/removal

Show what areas need to be focused on in the future

Personalized User Update Msgs

I think this idea needs some work.

Build a program (or POC demo) that scrapes and parses user reviews on TurboTax

Maintain some database/file that records usernames and complaints

Program emails or displays msgs to users addressing their complaints (new features that were introduced the next year, assistance, helpful links)