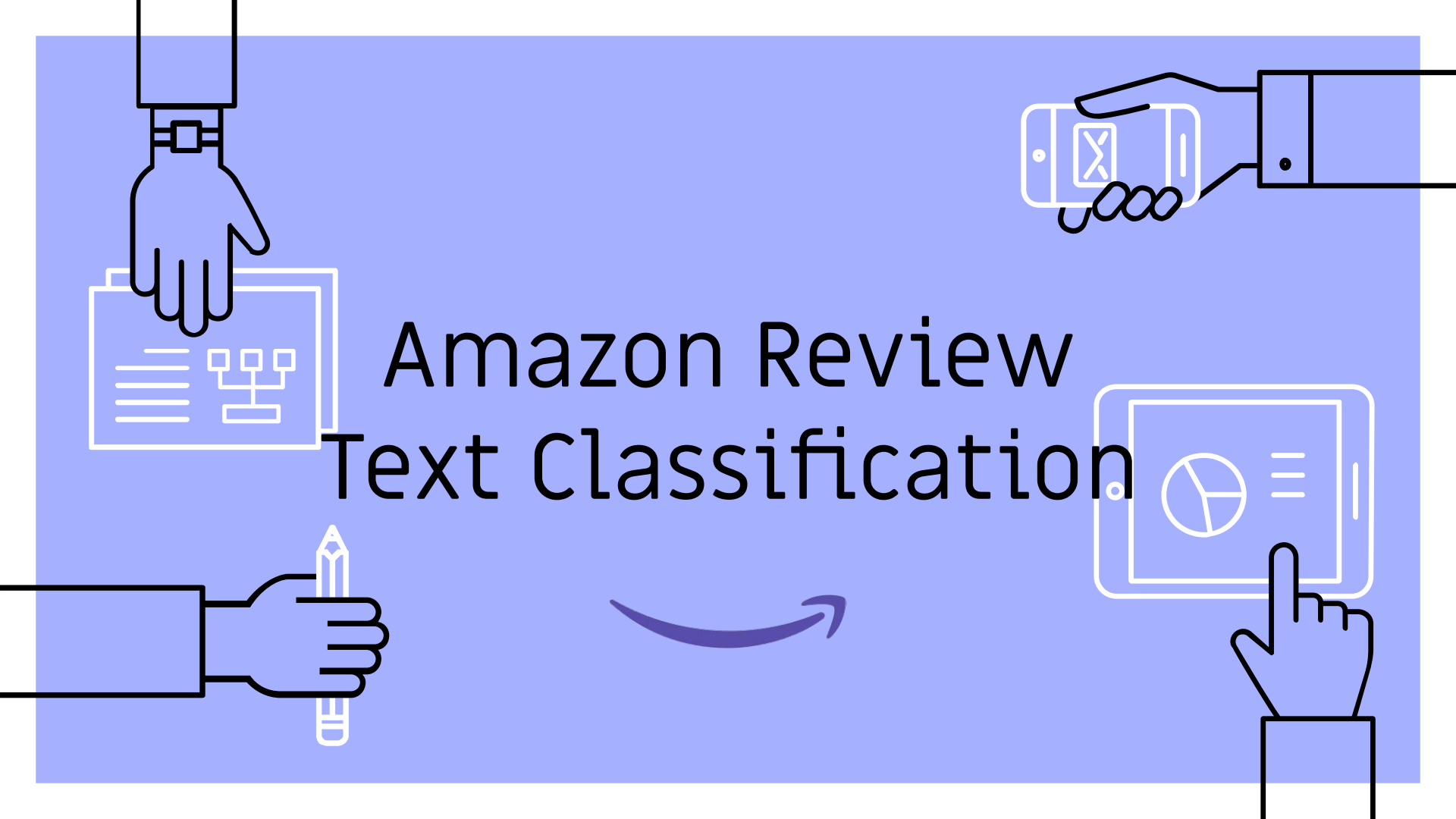


Amazon Review Text Classification

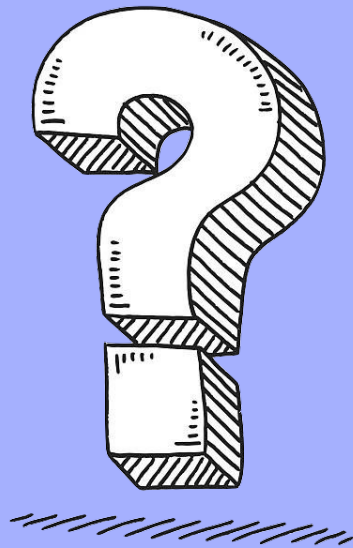


The Problem

Amazon sellers receive unlabeled textual data all the time

Emails, tweets, social media, company websites

Human classification is costly and surprisingly inaccurate



 **Crafting with Cat Hair: Cute Handicrafts to Make with Your Cat**
\$8.97  | FREE One-Day
Delivered tomorrow for FREE with qualifying orders over \$35. [Details](#) | [In Stock](#) [S](#)

★★★★★ **Worked like a charm.**

By [Rico](#) on July 12, 2014

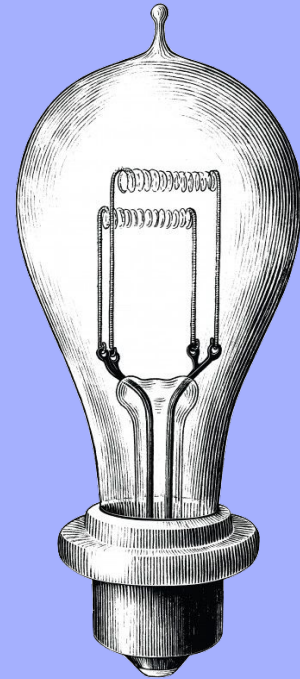
Format: Paperback

I purchased this book as I was tired of people sitting too near me on public transport.

[4 Comments](#)

272 people found this helpful. Was this review helpful to you?

[Report abuse](#)

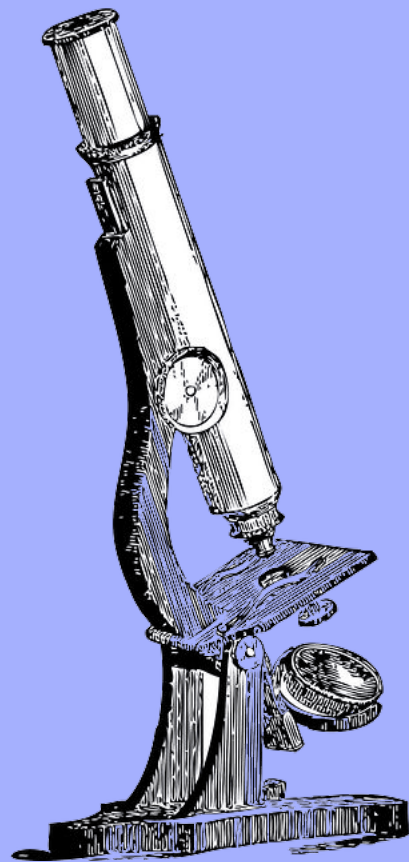


The Solution

Use labeled reviews to train machines

Use these machines to estimate
amazon star ratings

Gain valuable and reliable insight from
large amounts of textual data

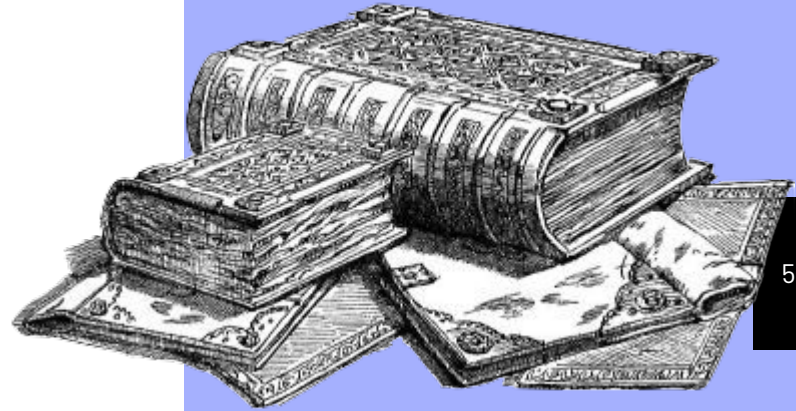


Handling Text Data

Remove common words that hold little meaning

Remove punctuation and numbers

Create numerical representations of words

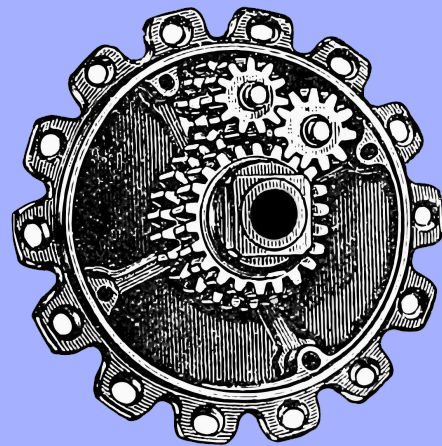


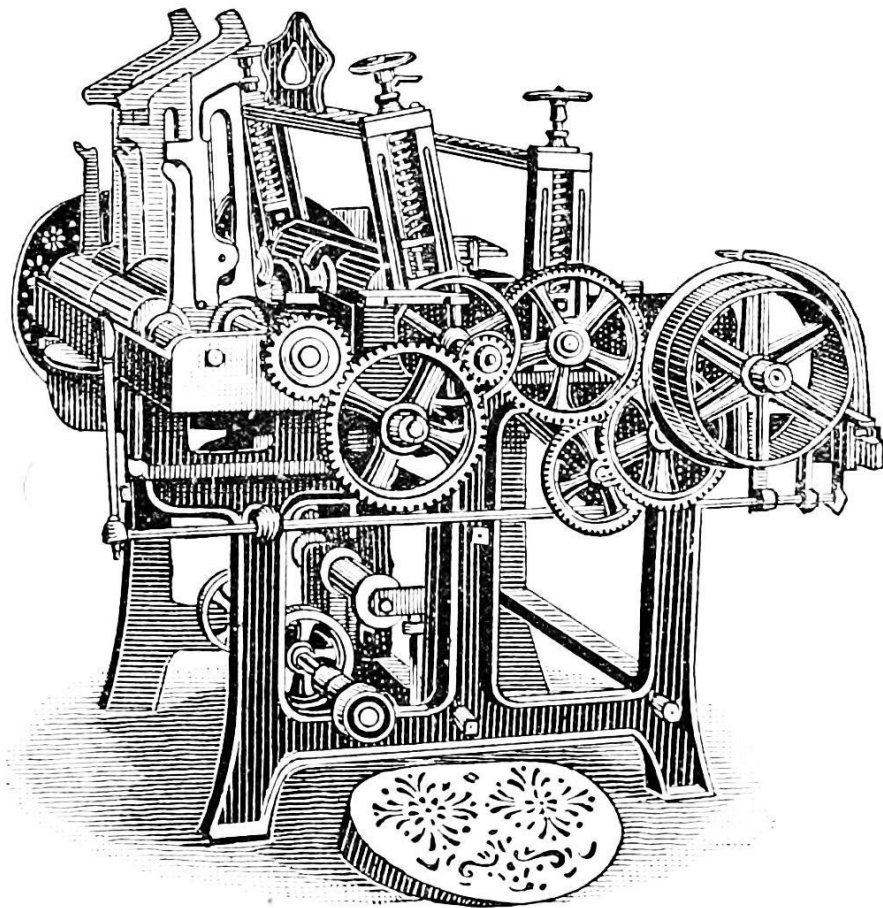
Modeling

Use the numerical representations to train a neural network

Include embeddings so that the numerical representations can improve

Use the model to make predictions using unlabeled data





The machine accurately predicts the correct class more than 66% of the time

When not predicting the correct class it almost always predicts an neighboring class





Moving Forward

Producing an app that can take in unlabeled feedback and make predictions

Use more data and more varied data to make more general predictions



THANKS!

Any questions?

You can find me at:
github.com/nismotie

