Proposal
Recommending Similar Dresses
Kseniya Kruchok

Problem:
Often when searching online for a new dress to buy, I don't have a solid idea of what exactly I am looking for. I might find one dress, that is almost what I want, but not totally. It would be great to be able to see similar items and have a wider choice of something related.

Client:
The solution can benefit both: large retail companies with many different dresses for sale (like Macy's, Bloomingdales, Nordstrom etc.) to implement or improve useful tool into their websites and mobile applications, and end users that can spend less time browsing and more time buying.

Data:
I have went through several fashion datasets:

1. MNIST Fashion Dataset https://github.com/zalandoresearch/fashion-mnist - it is an interesting alternative to traditional MNIST with handwritten digits, but gray scale images might not be enough to solve this problem
2. Categorization Dress Patterns from Open Data https://data.world/crowdflower/categorization-dress-patterns - over 15K dress images with the border around the dress, but only one category that is mostly about the color of the dress.
3. Kaggle dataset - https://www.kaggle.com/paramaggarwal/fashion-product-images-dataset#styles.csv - has about 30K dresses with several attributes: color, season, usage and general description
4. Deep Fashion Dataset http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html - a large collection (over 280K images) with about 1000 attributes. I will need to filter for dresses only

The latter dataset seems the most appropriate at the moment because of the large number of attributes. I will be starting with it and then keeping #2 and #3 as reserve

Solution:

1. Downloading images and filtering for dresses only
2. Exploring Available attributes and picking even amount of images in each category of interest
3. Performing transfer learning with pre-trained computer vision models
4. Trying a different approach of clustering images based on what we see on the image and not predefined attributes (word2vec approach)

Deliverables:

1. Code with visualizations and trained model
2. Presentation Slides
3. Github repository
4. Article