Silk Labs - Fashion Product Images NER Pipeline	
Contact	Alexander Sikand, asikand@bu.edu, 831-334-5344 Isaac Oshana, ioshana@scu.edu,
Organization	Silk Labs
Organization Description	Silk Labs is a company working to develop AI solutions for e-commerce and fashion brands.
Project Type	Natural Language Processing / Computer Vision
Project Description	The goal of this project is to take our internal dataset (25K+ product descriptions), create ground-truth labels to fine-tune a transformer-based NER model, and demonstrate the model's generalization to unseen data. We are also working to label product *images* and train a Mask-RCNN model to perform image segmentation, however generating ground-truth is much slower and time-consuming. If time allows, we will also attempt to make significant labeling progress for the images.
Data Sets	Proprietary dataset containing 25k+ images and product descriptions. Data mining, data munging, and data cleaning have already been performed by myself in recent weeks.
Suggested Steps	We have purchased two licenses of the labeling software prodi.gy that is created by the company behind the open-source library spaCy. The software runs offline on our own hardware and uses a model-in-the-loop to drastically increase the labeling speed. 1) Move the GoLang parallel web scraper built on top of Colly to serverless setup 2) Label ground-truth entities in product descriptions and titles 3) Train a custom NER model specific to the fashion industry 4) Evaluate the accuracy of the model 5) Serve the model in a production-like environment
Questions to be answered in Analysis	What kinds of entities can we detect with high accuracy (95%+)? How can we create a pipeline for merging the annotations of multiple labelers at once? Can we use the entities/attributes extracted from the product descriptions to increase the inference speed of our CNN models? Does the multi-modal data allow us to be a strong competitor against companies that aren't using multi-modal data? (We have looked at all publicly known competitors).