Daniells Edwards & SaweKrishna Gudipudi

24-787 Final Project Proposal

**Logistic Regression vs CNN for Jordan’s classification**

**Motivation**

Jordans are sneakers that are renown in the sneaker market due to the legacy left behind by the Basketball player icon, Michael Jordan. There are many different collaborations and models for these sneakers, but our project will be focused on the labeling of the sneakers collection from Jordan’s 1 – 23. Although there are some similarities in some of these models there are key features that differentiates one from the other and We would like to know if the model will be better represented by a logistic regression or a CNN.

**Data**

There has been attempts at creating an ML model for categorizing shoes from different brands, but these haven’t been as efficient due to the lack of datasets and the bias in brands’ websites. To create the dataset, we will write scrapers that will be able to download images from these sites and resizing them to a set ratio. To account for the bias in the shoes displayed by websites we will be focusing our model on just Jordan’s and making sure each label has relative the same size of training data. By using scrapers on different websites, we will be able to collect data from a wider pool which will help in varying the training set.

**Algorithm**

we found an article of somebody else attempting to create a sneaker classifier which will serve as a guide for when we get stuck on our implementation. Currently we are thinking of using a logistic regression model given that each of the sneakers’ labels have unique features that differentiates them from one another. For the CNN model we are thinking of building a model with just grayscale images without a reducing layer, and a different model taking the colors into account. Since the feature of the sneakers does not depend on the colors, we will be able to see if we can create a better model with a CNN without reducing its dimensions.

**Results**

The results will be collected by splitting the dataset into training and testing sets. We will be comparing the accuracy from CNN against the one for logistic regression with the model being trained with color and grayscale. By using both, the grayscale and the full color data, we will be able to tell which model was more efficient at finding a trend across the different labels of the sneakers.

**References**

1. [ML model example](https://medium.com/@ryxnszn/sneaker-machine-learning-ac587bbbe92)
2. [Jordan Collection](https://www.jordan.com/collection/air-jordan-14)