Categorizing Images in the Fashion MINST Dataset Using Computer Vision and Machine Learning

Bryce Anthony and **Christos Koumpotis**
{branthony,chkoumpotis}@davidson.edu
Davidson College
Davidson, NC 28035
U.S.A.

Abstract

In this paper, we implement 6 different machine learning models to categorize the images in the Fashion MNIST DATASET. We used a Random Forest Classifier model, a K-nearest neighbors model, and a C-Support Vector Classification model using a support vector machine with a Radial Basis Function as the kernel. After testing our model we preprocessed the dataset using binarization to reduce the noise in the dataset which ultimately lowered the validation accuracy of each model. We found the most accurate model to be the random forest classifier using non-augmented images which performed nearly 10x better than the dummy model.

1 Introduction

Computer vision is a field of artificial intelligence that trains computers to interpret visual images and unlocks the potential for computers to understand the visual world. As technology advances and the use of artificial intelligence continues to be adopted by a variety of industries, the use of artificial intelligence and computer vision for fashion could be an especially interesting domain, especially with the prevalence of fashion e-commerce.

The Fashion MNIST dataset consists of 60,000 black and white images of articles of clothing that belong to 10 different categories. The categories of clothing present in the dataset are: T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle-boot. Our challenge was to create a machinelearning model capable of classifying the clothing images into their respective categories based on the pixel values of each image. To accomplish this we built on existing research from (K V and K. 2019), (Greeshma and Sreekumar 2019), and (Xiao, Rasul, and Vollgraf 2017) all of whom tackled the same problem of correctly classifying images from the Fashion MNIST dataset. We preprocessed our images using binarization and used support vector machine models based on the comments made by Greeshma and Sreekumar in their paper.

In the remainder of the paper, we first provide some background information on the machine learning algorithms we implemented, the details of our experimental approach, and discuss our results as well as the broader impacts of our research. lastly, some concluding remarks will be made with some recommendations for future research.

2 Background

Despite the advancements in computer vision there is no empirically defined best way to work with image data. This is due to the way that images are represented on computers. For this task the clothing images were represented in an array with the dimensions 1x784, where each item in the array represents a pixel of the image and each value in the array took on a value between of 0 and 255. This array representation of an image makes the task of classifying images much different than humans are used to as no one to one correlation exists between the way computers interpret images and the way humans do.

In addition to the format of the images, the representation of color on a scale from 0 to 255 adds complexity to the task. In order to decrease the complexity of the task and boost model performance, we got the idea to eliminate the concept of color in the images and represent the pixels in each image by setting each non-zero pixel value to 1; which we refer to as binarization.

3 Experiments

To select our best model we trained several models and used the accuracy of each model on its validation set to evaluate its quality. In terms of the data used in our experiments, we had a train dataset with 60,000 samples and a test dataset with 10,000 samples with each class being equally represented in the train and test datasets.

The 10,000 observations in the test dataset were set aside and reserved for evaluating our final model, while 60,000 observations in the train dataset were used to train each of the models. To train each model we did a 5 fold cross validation where 12,000 observations were randomly selected to be used as validation data and the other 48,000 observations were used as training data for each of the 5 folds. The 12,000 observations from each fold were reserved and after the model was trained on all 5 folds, a dataset consisting of the 60,000 reserved observations ($12,000$ observations * 5 folds) was used to validate the accuracy of the model.

We tested 3 different models: a model implementing a k-nearest neighbors classifier, a model implementing a random forest classifier, and a model implementing a support

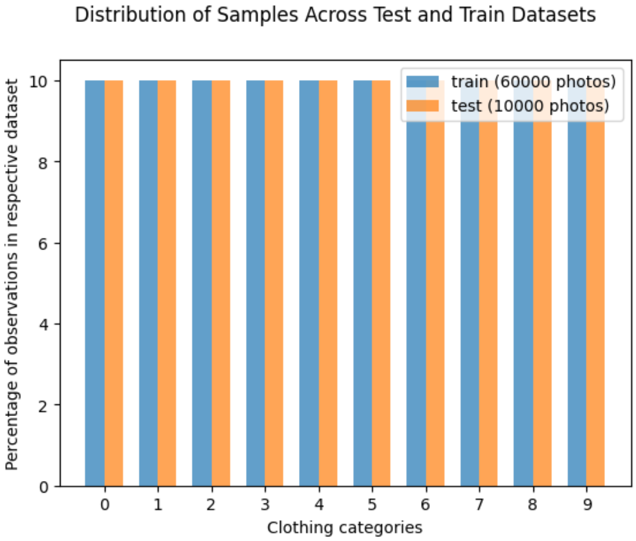
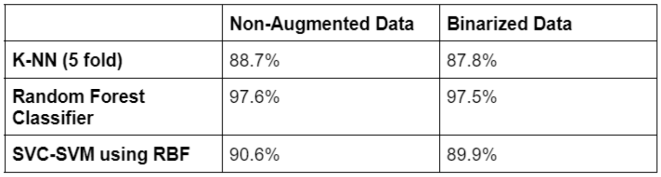


Figure 1: Distribution of samples across test and train datasets

vector machine (SVM) with a radial basis function as the kernel. Each of these 3 models was evaluated based on their performance using the image data directly from the dataset and using a binarized version of the images from the dataset which resulted in validation scores for 6 distinct models.

4 Results

Based on the validation mean accuracies of our models; the Random Forest Classifier trained on non-augmented images, was the most accurate with a validation accuracy of 97.6%. The accuracy for the Random Forest Classifier was about 7 percentage points higher than our next best-performing model and roughly 87 percentage points higher than the dummy model which had an accuracy of 10%. Our other two models trained on non-augmented images, were a K-nearest neighbors model and a C-Support Vector Classification model using a support vector machine with a Radial Basis Function as the kernel. These models had accuracies of 88.7% and 90.6% respectively. We hypothesized that the lower accuracy of the K-NN model is due to the combination of the way the images are represented and the way the K-nearest neighbors model works.



	Non-Augmented Data	**Binarized Data**
K-NN (5 fold)	88.7%	87.8%
Random Forest Classifier	97.6%	97.5%
SVC-SVM using RBF	90.6%	89.9%

Figure 2: Accuracy of the 3 models during validation on the non-augmented data and the binarized data

Because images are represented as arrays of integers with values ranging from 0 to 255 depending on the color of each pixel, and the K-NN model tries to map all the data points (images) according to their pixel values, in a multidimensional graph and identify the closest neighbors of any given test sample. Therefore, shades play an important role in plotting items in the graph and thus two images depicting the same type of clothing, might appear further away from each other, while two images of different items might appear closer to each other, because of the different values that pixels take according to their shades.

We tested our hypothesis by preprocessing our data using the binarization process. We thought this would allow the models to focus their predictions on the shape of the items rather than their coloring. After running the models with the binarized data we noticed that the validation accuracy of each model had decreased by about a percentage point meaning our hypothesis was rejected. Although doing the binarization process on our image data didn't lead to better model performance, the slight decrease in the accuracy of the models suggests that the colors in the images may contain both important information along with some noise.

The accuracy of the random forest classifier could be attributed to its sampling process as it considers a greater variety of the training data to make a prediction which makes the model less prone to making bad predictions based on the variations in color and/or slight differences in the shapes of clothing items. The SVM model had a similar accuracy to the K-NN model, as using the RBF kernel focuses on the distance between points to determine their similarity. At the same time, it presented a slightly better accuracy, which we believe is because it doesn't limit itself to a specific number of neighbors to make a prediction, and therefore is less influenced by inaccuracies due to coloring or slight differences.

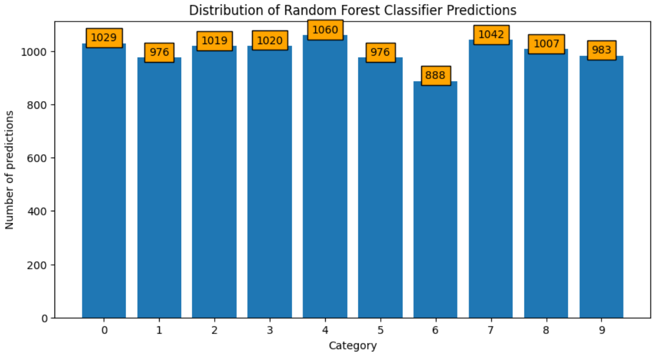


Figure 3: Distribution of classifications from the Random forest classifier model- adapted from (Greeshma and Sreekumar 2019)

To test our best model, we used the 10,000 sample test dataset and computed predictions for each of the images. The predictions depicted in the image below were 86.95% accurate which was notably about 10 percentage points worse than the model did on the validation dataset. We believe this was due to the model overfitting the validation dataset and leave this as a point for further research. Another important consideration in evaluating the model is the similarities between different types of clothing items. Despite there being 10 different categories for clothing, some of the categories and clothing items are hard to distinguish between even for humans. Overall, the results from the test dataset suggests that the model is generally effective at predicting the clothing category of images in the dataset.

5 Broader Impacts

Using machine learning to identify pieces of clothing and accessories can have a few potential issues. First of all, in our case, the training data includes clothes and accessories that are prevalent mostly in western cultures. This not only makes the model extremely inaccurate and ineffective in identifying items from more diverse cultures that include different types and versions of clothing and accessories but also reinforces western biases. More precisely, if such a model was to be widely used, it would reinforce and impose the culture of the West in all of its implementations, which would further perpetuate the discrimination cultural minorities face not only in the fashion industry but in society as a whole. At the same time, clothing and fashion are a form of self-expression that fosters and encourages innovation and creativity, and could even be described as a form of art. Therefore classifying a form of art into some distinct and common categories would result in significantly reducing the applicability of the model and depending on the implementation of the model, could also result in a form of suppression of creativity in the fashion scene or even to an individual level.

Another important consideration for models like ours has to do with the nature of the fashion industry. More elaborately, the fashion industry is very fast-paced and evolving, therefore, the model is prone to becoming obsolete and inaccurate very fast. In other words, unless the model is retrained and updated frequently and accurately to reflect the newest evolutions of the fashion scene, it will end up reinforcing old ideas, and trends. Lastly, fashion is evolving to break the barriers of classifications, such as t-shirts being worn as dresses, and therefore an accurate model that allows for creativity and self-expression would need to take into account the way that items are being used and not just what they appear to be.

6 Conclusions

We created 2 variations of 3 different models to predict the clothing category of images in the Fashion MNIST dataset. Images were represented as arrays with each value in the array representing a pixel of the image. Our most accurate model was the random forest classifier using non-augmented input data and this model had an accuracy of 86.95% on the test dataset. We also trained a Support Vector Machine model using a Radial Basis Function kernel and a K-nearest neighbor model which were both less accurate than the random forest classifier model. We hypothesized that the variation in pixel values affects the accuracy of the latter two models because of their focus on the distance between points in computing predictions. This led us to binarizing our images before feeding them to the models which proved to be ineffective and resulted in all three of our models being less accurate.

For future research we recommend that a more complex preprocessing of the data be implemented as we believe that a different approach to preprocessing the data could reduce the impact of shading and focus on the shape of the items in the dataset leading to a more accurate model. We suggest that normalizing the data could be a more effective way of doing so.

7 Contributions

Bryce and Christos split up the coding part of the assignment Bryce Coded the K-NN and Random forest classifiers while Christos implemented the Svc-Svm model with the RBF kernel and coded the binarized data set. For the writing of the paper Bryce wrote the introduction, and experiments, references, and edited other parts of the paper. Christos wrote the Background, Results, Broader Impacts and Conclusion. Bryce created figure 3 and Christos created figure 2, we worked together on figure 1.

References

Greeshma, K., and Sreekumar, K. 2019. Hyperparameter optimization and regularization on fashion-mnist classification. *International Journal of Recent Technology and Engineering (IJRTE)* 8(2):3713–3719.

K V, G., and K., S. 2019. Fashion-mnist classification based on hog feature descriptor using svm. *International Journal of Innovative Technology and Exploring Engineering* 8:960–962.

Xiao, H.; Rasul, K.; and Vollgraf, R. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms.