**Abstract for CIND860 Advanced Data Analytics Project**

**By: Hasib Chowdhury (500439135)**

**Supervisor: Dr. Tamer Abdou**

**Date of Submission: September 25, 2023**

For this project, I will use the Fashion MNIST dataset (<https://www.kaggle.com/datasets/zalando-research/fashionmnist>) from the Kaggle website to conduct the research to fulfill the “Advanced Data Analytics Project (CIND860)” requirements.

There are a total of 70,000 images for this dataset, which is split 60,000 for training and 10,000 for testing (comprising of 28x28 grayscale images). This exact dataset is also built-in in the keras library, which shows the exact images (shirts, pants, sneakers, etc.) in its pixel grayscale form as this is not available on Kaggle. The Kaggle dataset only has 2 .csv files for training and testing for the 784 pixels in the dataset (ranging from 1 to 255 in darkness of the image).

The theme that has been chosen for this project is the deep learning theme (specifically image classification on various Fashion attires) and the technique being used is Convolutional Neural Network (CNN) on the Fashion MNIST dataset.

Some of the research questions this project will go into detail is what models of the CNN is the most efficient to use. I would also go into how this particular dataset set the foundations for benchmarking machine learning algorithms in general and from the original MNIST dataset. *“Fashion businesses in general have used CNN on their e-commerce platforms to solve many problems such as clothes recognition, clothes search and recommendation. A core step for*

*all of these implementations is image classification. However, clothes classification is a challenge task as clothes have many properties, and the depth of clothes categorization is highly*

*complicated.” [1].*

This project will look into various properties for the clothes such as which sizes are used (small, medium, large, XL) in certain attires as well as how different attires suit different types of people. This project will investigate is if there is a discrepancy in the CNN model when it comes to different age and genders like different fashion attire such as T-shirs/tops for Men and Women compared to kids or teenagers*.* Lastly, this project will see how the CNN compares with the SVM as mentioned in ***other*** research papers and other types of models when it comes to accuracy and precision of the dataset, as well as data labelling of different types of attire.

I will use Python as the main programming language. I will also look at which specific models of the CNN architecture are the most commonly used when evaluating the Fashion MNIST dataset in deep learning models. Throughout the fashion industry and also in Fashion e-commerce and in online retail such as Amazon and E-bay, the market has been growing in recent years and the “CNN model in particular has been shown greater efficiency in image c1assification” [3]. This is what this project will look into as the main technique that is going to be used throughout the project.

References

[1] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

[2] Kayed, M., Anter, A., & Mohamed, H. (2020, February). Classification of garments from fashion MNIST dataset using CNN LeNet-5 architecture. In *2020 international conference on innovative trends in communication and computer engineering (ITCE)* (pp. 238-243). IEEE.

[3] LEITHARDT, V. (2021). Classifying garments from fashion-MNIST dataset through CNNs. *Advances in Science, Technology and Engineering Systems Journal*, *6*(1), 989-994.