**Declaration**

We hereby declare that the project work is entitled “*Image Scrapper In Bash”* has been prepared by us during the Odd Semester 2017-2018 under the guidance of Ms. Deepthi L and Ms. Sangeetha S H, Department of Information Technology.

We also declare that this project is the outcome of our own efforts We also confirm that, the report is only prepared for our academic requirement not for any other purpose. The information incorporated in this project is true and original to the best of our knowledge.

Moksh Jain - 16IT221

Suyash Ghuge - 16IT114

Abhishek Kamal - 16IT201

Nishanth Hebbar - 16IT234

**Abstract**

Machine Learning has become one of the most popular fields of Computer Science. The rapid advancements are allowing Machine Learning Models to learn more and more complex tasks. The progress in the field of Computer Vision though has been exponential. Various Image Classification models have state-of- the-art accuracy and have lower error rates than even humans. This tremendous progress has been powered by the large datasets, on which the models are trained. These datasets are huge and take hundreds of man hours to properly compile and label. So for someone just starting off, preparing a new dataset in a big task. But, since this is the age of the Internet, there a vast amount of data ready to be used. This project enables the user to do just that. The Image Scraper is to be provided with a list of categories to scrape for in a file, the number of images to scrape, and the dimensions of the output images. The Image Scraper then fetches results from the Google Image Search for each of the queries, using curl or wget, depending on the user’s choice. We then capture the links to all the images returned for the search query using pattern matching using egrep and awk and save them to a file for later use. Once the list is complete we download each of the image and save them to a folder corresponding to the query. Once downloaded, the image is resized using the convert command, from the ImageMagick package. The image is resized, since for training as well as testing a Machine Learning Model expects all the images to be of fixed dimensions. At the end, the results are stored in a folder, with the images for each of the query in a separate subfolder. This dataset can then very easily be imported in Python, which is the most widely used programming language for Machine Learning. In summary, the Image Scraper allows the user to build a toy dataset for learning as well as experimenting with Machine Learning.

**Table of Contents**

1. Introduction

1.1. Challenges 1

1.2. Motivation for the Work 2

2. Methodology and Framework

2.1. System Requirement 3

2.2. Algorithms, Techniques 4

3. Implementation

3.1. List of Main Unix Commands 5

3.2. Uses and its Syntax 6

4. Result and Analysis 8

5. Conclusion 10

References 11

**List of Figures**

Figure 4.1

* Output of the Script 8

Figure 4.2

* Downloaded Folders 8

Figure 4.3

* Downloaded Files 8