**Title: Fake Brand Logo Detection**

**Introduction:**

Fake brand logos, also known as counterfeit or knockoff logos, are logos that are intentionally designed to look like the logos of well-known brands but are not authorized or produced by the brand owner. These logos are often used to sell fake or low-quality products to unsuspecting consumers, which can harm the reputation of the brand and the safety of the consumer.

The problem of fake brand logos is widespread and can affect a wide range of industries, from fashion and luxury goods to electronics and pharmaceuticals. In fact, the global value of counterfeit goods is estimated to be in the hundreds of billions of dollars annually.

**Objective**:

The main objective of this project is to develop a machine learning-based model by using API that will connect to the website that can accurately detect fake brand logos and distinguish them from genuine ones, with a focus on copyright infringement.

**Existing Project:**

There is research done by SRM Institute of Science and Technology in which they have used correlation method and Tesseract-OCR to detect fake logos. Tesseract-OCR is a tool that can recognize text in images. It works well, but sometimes has trouble with unique fonts due to which they have 80% accuracy. After thinking about doing a project on same topic and when we were thinking how we can increase the accuracy, we came across beer logo detection project done by Meerkat Cv, a computer vision tech company specialized in face recognition and object detection. In their project they have used machine learning. They have trained their system to detect logos of the following beer brands: Budweiser, Bud Light, Corona, Guinness, Heineken and Stella Artois on social media. After going through their project we thought to use machine learning in detecting fake brand logos keeping in mind that machine learning algorithm works well with unique fonts as compared to Tesseract-OCR. We will be using machine learning algorithms to increase accuracy by more than 80%.

**Dataset:**

Our dataset consists of 50 brands with 100 original images and 20 fake images of each brand. We analyzed the logos2k+ and Flicks dataset. In these two datasets there are images of only original logos. Whereas for our project we needed both fake and original images. We didn’t have any proper dataset according to our requirements. So, we created our own dataset. For original logos we downloaded the images from every brand’s official site and from google. While in case of fake logos ,some brands fake logos were easily available on google and Pinterest so we downloaded from these sites and for brands whose fake logos were not available ,we generated the logos .We generated the fake logos using “Brand crowd” website and PowerPoint. All the downloaded images were in different format for e.g. Some images were in JPEG while some were in PNG format. So, by using PIP library from python, we converted format of every image to JPG. After this we renamed all the images as some pattern.

**Five c’s**

**Consent-** We ensure proper consent throughout the progress of our project. We took several steps to ensure consent for our project. We checked the terms and conditions of the websites to use the images when we were collecting images from the websites. We clearly explained information about our project, i.e., the purpose of our project and how we plan to use the dataset in our objectives. And ensure that the images used in the project are kept protected and confidential i.e. we will use the images only for our project purpose and not for any other unintended purpose.

**Clarity**- The purpose of the project is to detect fake brand logos. The website that we are developing will be designed only to identify fake logos and prevent their use and not be used to infringe on the rights of legitimate brands. The criteria for identifying fake logos will be based on the similarity of the logo to the original.

**Consequences**- Consequences is an important part of detecting fake brand logos, it makes companies and consumers protected from fraud and any misleading data. Our dataset is private and there is no chance that our dataset is misused. It also makes the companies make sure that the data is safe.

**Control**- We ensure that the project is carried out in a responsible and ethical manner. We ensure the dataset is used only for its intended purpose and is not misused for any other purpose. we will make sure that only our group members and professor have access to the dataset as it has fake logos of brand which our generated by us and misusing them can harm brands reputation. We will have a data-sharing agreement when we share files with third parties.

**Consistency**- Consistency is an important aspect of fake brand logo detection. Our dataset contains fake and original images of different brands, where each image with the same format and unique name. It’s very easy to identify each brand in the dataset. By using this model can be trained easily and processed the results are accurate and reliable.

**Our project-**

We have created target\_dict to map labels of the dataset ("Original" and "Fake") to integer labels (0 and 1). We loaded all the images as grayscale image and using smart\_resize() all the images are resized to 128x128 pixel size .We have used train\_test\_split() from scikit-learn library to split the images and labels into training and testing sets(20% of the data is used for testing and 80% for training)we have developed a CNN model using the Keras API in TensorFlow to classify images of brand logos as original or fake. The model architecture consisted of 6 convolutional ,2 max-pooling layers, 1 flatten layer, 2 dense layers along with 2 dropout and 2 batch normalization layers to prevent overfitting. The dataset was preprocessed to ensure that the images are of consistent size. The training dataset consisted of images of both authentic and fake brand logos, while the validation and test datasets were used to evaluate the model's performance. The model is trained for 20 epochs with a batch size of 25, using the RMSprop optimizer with categorical cross-entropy loss function. The model's accuracy is used as the primary metric to evaluate its performance on the validation and test datasets. The results showed that the model was able to accurately classify images of brand logos into their respective categories, with a test accuracy and validation accuracy of **0.8633333444595337** and **0.8550000190734863** respectively. We have created two plots to visualize the performance of the model during training. The first plot shows the training and validation accuracy per epoch. The second plot shows the training and validation loss per epoch.