Quality Inspection Model

Presented To: **Dr. Mohammad Awde**

Presented By: **Hassan Abbas-5615**

Ahmad Ismail-5476

2020-2021

Table of Contents:

I. Introduction

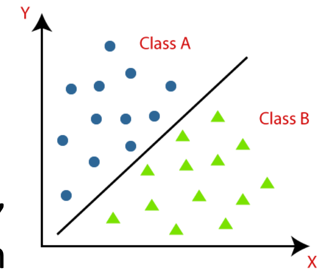
Casting is a manufacturing process used by the forging industry to produce an object (often metal) of a specific desired shape. This is achieved by pouring hot molten metal into a mold that contains a hollow cavity of the exact required shape. As the metal solidifies it acquires the shape of the hollow cavity.

Casting defect is an undesired irregularity in a metal casting process. There are many types of defects in casting like blow holes, pinholes, misruns, burr, shrinkage defects, mold material defects, pouring metal defects, metallurgical defects, etc. Defects are an unwanted thing in casting industry.

To detect such defective products, the quality inspection model was proposed based on machine learning methods and specifically on the KNN classification algorithm. This model worked with an augmented dataset that contain a total of 7348 images where all images are 300*300 pixels grey-scaled images. The main aim of this model is to classify a product as either OK or Defective. The dataset is already split into training defective has 3758 and OK has 2875 images) and testing (defective has 453 and OK has 262 images)

II. What is machine learning?

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs** that can access data and use it to learn for themselves.



The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is to allow the computers learn automatically** without human intervention or assistance and adjust actions accordingly. (What is the Definition of Machine Learning? | Expert.ai, 2021)

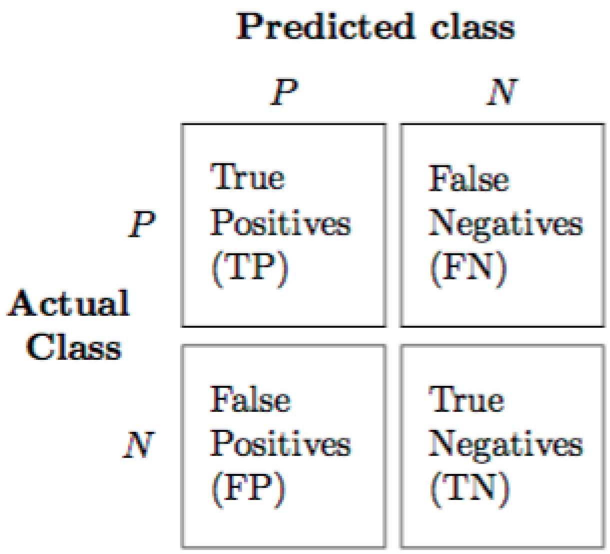
III. Classification algorithm:

Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y). (Asiri, 2018)

For example, spam detection in email service providers can be identified as a classification problem. This is s binary classification since there are only 2 classes as spam and not spam. A classifier utilizes some training data to understand how given input variables relate to the class. In this case, known spam and non-spam emails have to be used as the training data. When the classifier is trained accurately, it can be used to detect an unknown email.

Classification belongs to the category of supervised learning where the targets also provided with the input data. There are many applications in classification in many domains such as in credit approval, medical diagnosis, target marketing etc.

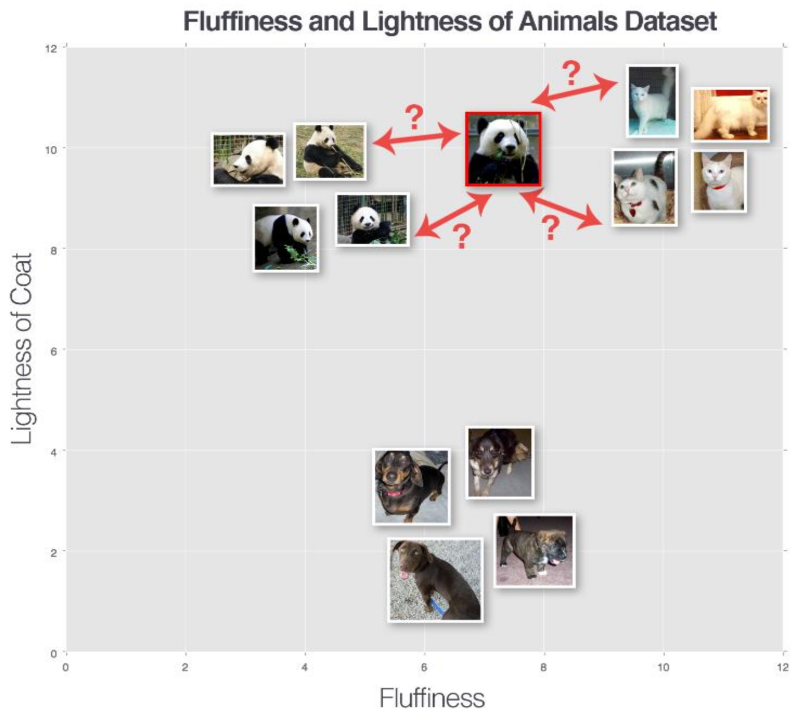
A good way to visualize binary classification is using the confusion matrix:



IV. K-Nearest-Neighbors (KNN) Classifier:

The k-Nearest Neighbor classifier is by far the simplest machine learning and image classification algorithm. In fact, it's so simple that it doesn't actually "learn" anything. Instead, this algorithm directly relies on the distance between feature vectors.

Simply put, the k-NN algorithm classifies unknown data points by finding the most common class among the k closest examples. Each data point in the k closest data points casts a vote, and the category with the highest number of votes wins. (Rosebrock, 2017)



In order for the k-NN algorithm to work, it makes the primary assumption that images with similar visual contents lie close together in an n-dimensional space. In the figure below we can see three categories of images, denoted as dogs, cats, and pandas, respectively. In this pretend example we have plotted the "fluffiness" of the animal's coat along the x axis and the "lightness" of the coat along the y-axis. Each of the animal data points are grouped relatively close together in our n-dimensional space. This implies that the distance between two cat images is much smaller than the distance between a cat and a dog.

However, in order to apply the k-NN classifier, we first need to select a distance metric or similarity function. A common choice includes the Euclidean distance:

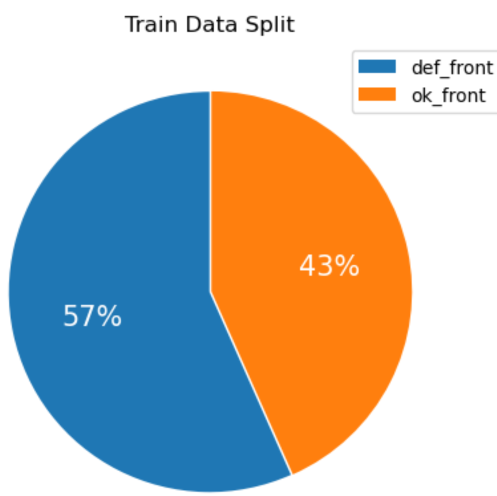
$$d(p,q) = \sqrt{\sum_{i=1}^{N} (q_i - p_i)^2}$$

However, other distance metrics such as the Manhattan/city block (often called the L1-distance) can be used as well:

$$d(p,q) = \sum_{i=1}^{N} |q_i - p_i|$$

V. The Quality Inspection Model

This project was completely implemented in python, we started our model by importing the necessary packages, like pandas, numpy, opencv, sklearn and other packages that helped us parsing files and displaying outputs. The next step was loading our dataset and visualizing its distribution.



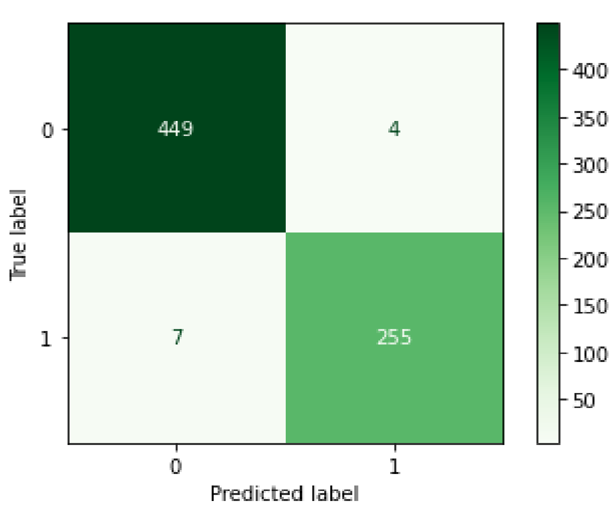
Here we can notice that there is a slight imbalance in the dataset towards the defective products, this imbalance won't have an effect on our model since it's relatively small. Next, we compiled the images in the dataset by resizing and converting them into numpy arrays, as follows:

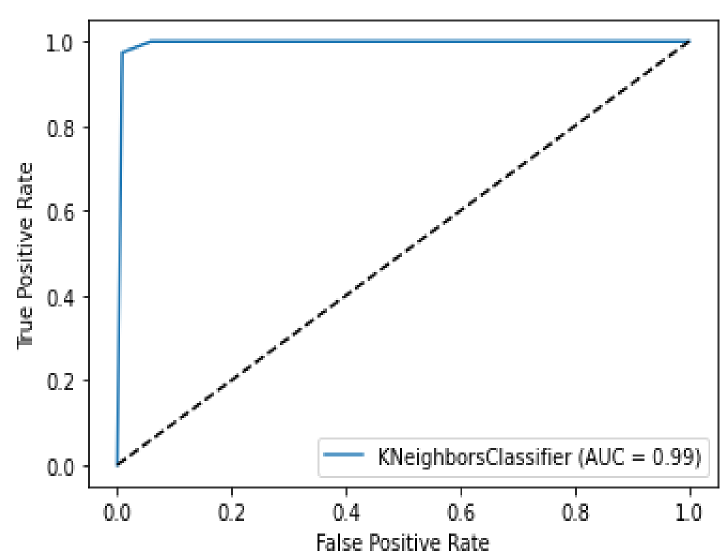
```
image_data_train = []
image_target_train = []

for title in files:
    os.chdir('C:\\Users\\user\\Documents\\Python Scripts\\New folder\\casting_data\\casting_data\\train\\{}'.format(title))
    counter = 0
    for i in data_train[title]:
        img = cv2.imread(i,0)
        image_data_train.append(cv2.resize(img,(width, height)).flatten())
        image_target_train.append(title)
        counter += 1
        if counter == sample_size:
            break
    clear_output(wait=True)
    print("Compiled Class",title)
image_data_train = np.array(image_data_train)
size_train = image_data_train.shape[0]
image_data_train.shape

train_images = image_data_train / 255.0
test_images = image_data_test / 255.0

train_labels = labels.transform(image_target_train)
test_labels = labels.transform(image_target_test)
```

The ‘width’ and ‘height’ are variables pre-defined to 15 pixels at the beginning of the code, this number was firstly chosen arbitrary and fine-tunned later. We also divided the dataset by 255, which is the range of color in pixels, in order to normalize the data.

The last step before prediction is to fit our data in the KNN model with a distance metric of 2 and calculate the final classification accuracy using our test dataset and plotting our results:

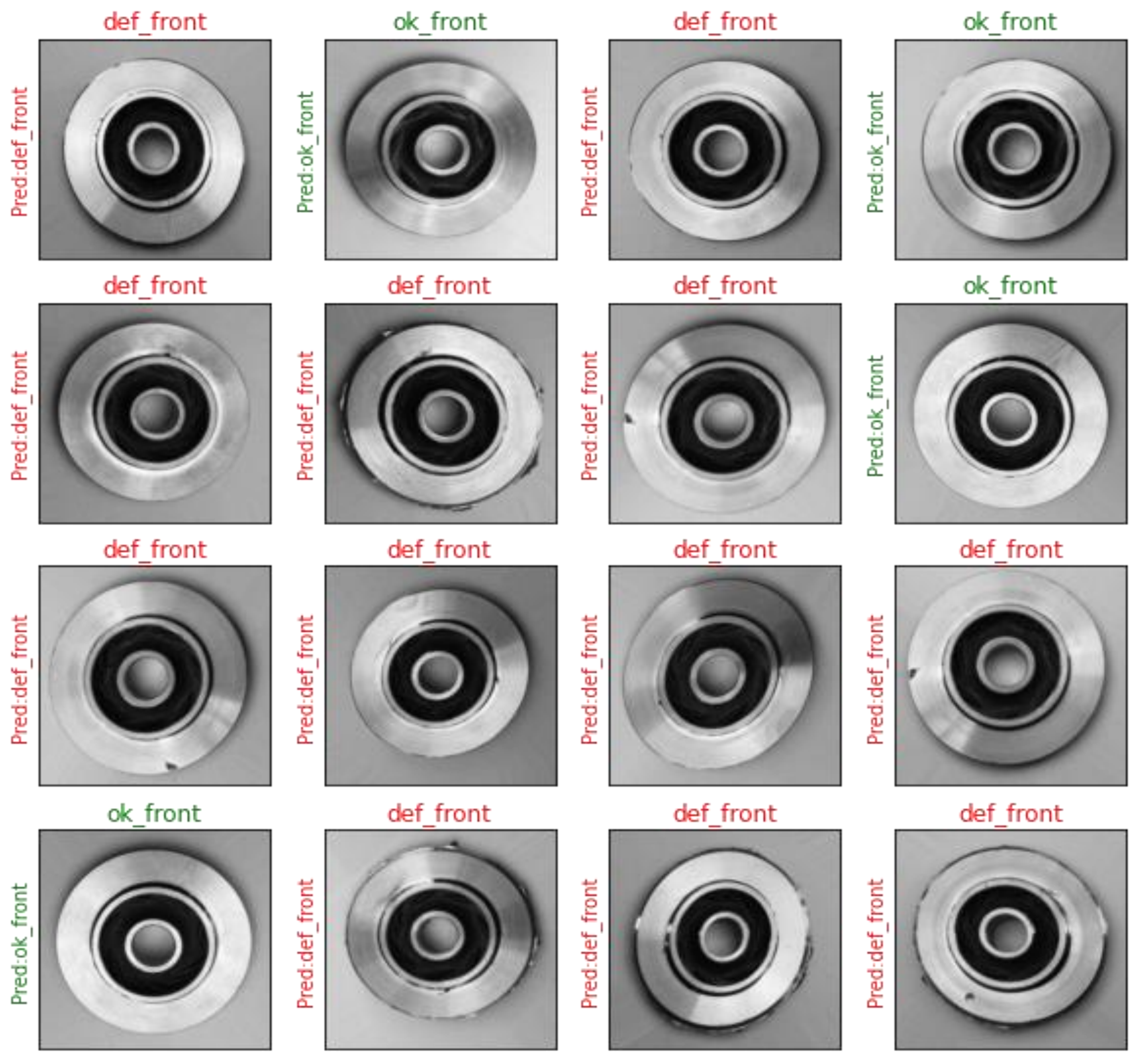
Acc: 0.98





We obtain an accuracy of 98% classification accuracy, which is an extraordinary number for such a simple algorithm.

The last step left, is to test the classifier on randomly selected images, we defined a loader prediction function, and fed the images to the model yielding uplifting results.



VI. Conclusion

Although the quality inspection model is simple and easy to implement, it still managed to provide us very good results. Such model can be implemented in the forging industries and thus remove their use for a whole quality inspection department, instead they could make use of this model to detect defects that are sometimes hard for the human eye to notice.

VII. References

1. Expert.ai. 2021. *What is the Definition of Machine Learning? | Expert.ai*. [online] Available at: <https://www.expert.ai/blog/machine-learning-definition/>
2. Asiri, S., 2018. *Machine Learning Classifiers*. [online] Medium. Available at: <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>
3. Rosebrock, A., 2017. *Deep learning for computer vision with Python*.