

# USING CONVOLUTIONAL NEURAL NETWORK TO PREDICT LOAN REPAYMENT VIA SELFIES.

Deep Learning IndabaX Nigeria 2021.  
Precious Kolawole, Lendsqr NG.



## Abstract

Repayment is the core of credit without it every lending business model falls apart. Lenders spend a lot of time on data, algorithms, and countless strategies to model the propensity of a prospective borrower to pay back. But this approach is difficult in Nigeria because of poor data. However, with the advent of Convolutional Neural Networks (CNN) we dived into the possibilities of applying ML to predict borrowers who would default using their selfies.

## Introduction

Previous and current methods in the industry leveraged the use of machine learning algorithms on transactional data to predict repayment. However, computer vision has played a huge role in facial sentiment analysis and identifying different facial characteristics, hence, we proposed a Convolutional Neural Network (CNN) model to predict loan repayment using borrower's selfies (pictures), by extracting useful information and catching complex unseen features from the images. Our first experimental result has helped us to look out for potential borrowers that'd not default.



Samples shows selfies of Borrowers(images have been masked for protection of Identity).

## Data

- Selfies were obtained from Database stored in AWS Cloud.
- Selfies were cropped to 240 by 240 pixels with a black colored padded background.
- Selfies were labelled into two classes; **Settled** and **Past Due**.
- Training data contains 2200 of the first class, and 1800 of the second class. This was further splitted into train, validation and test set.

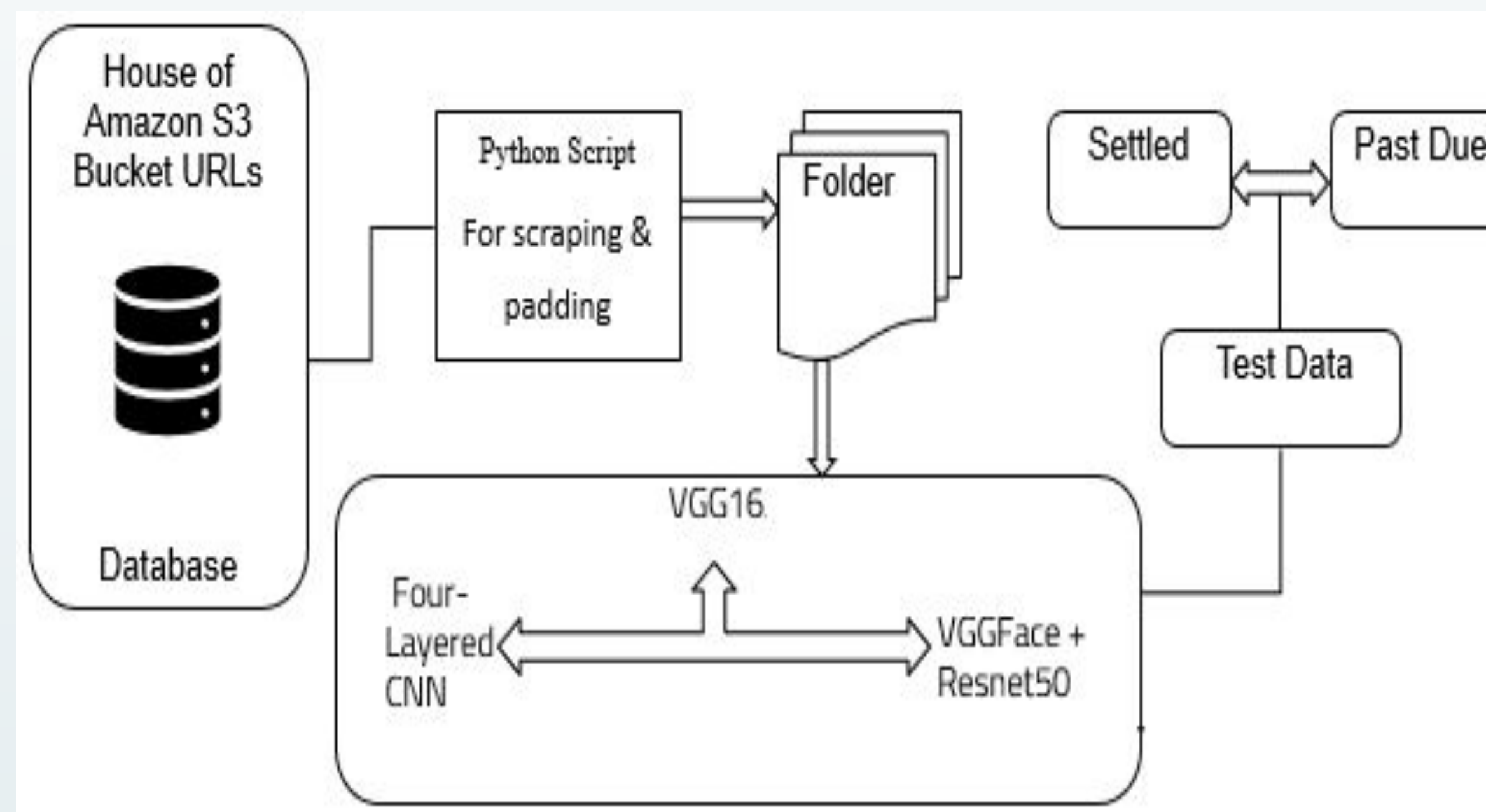
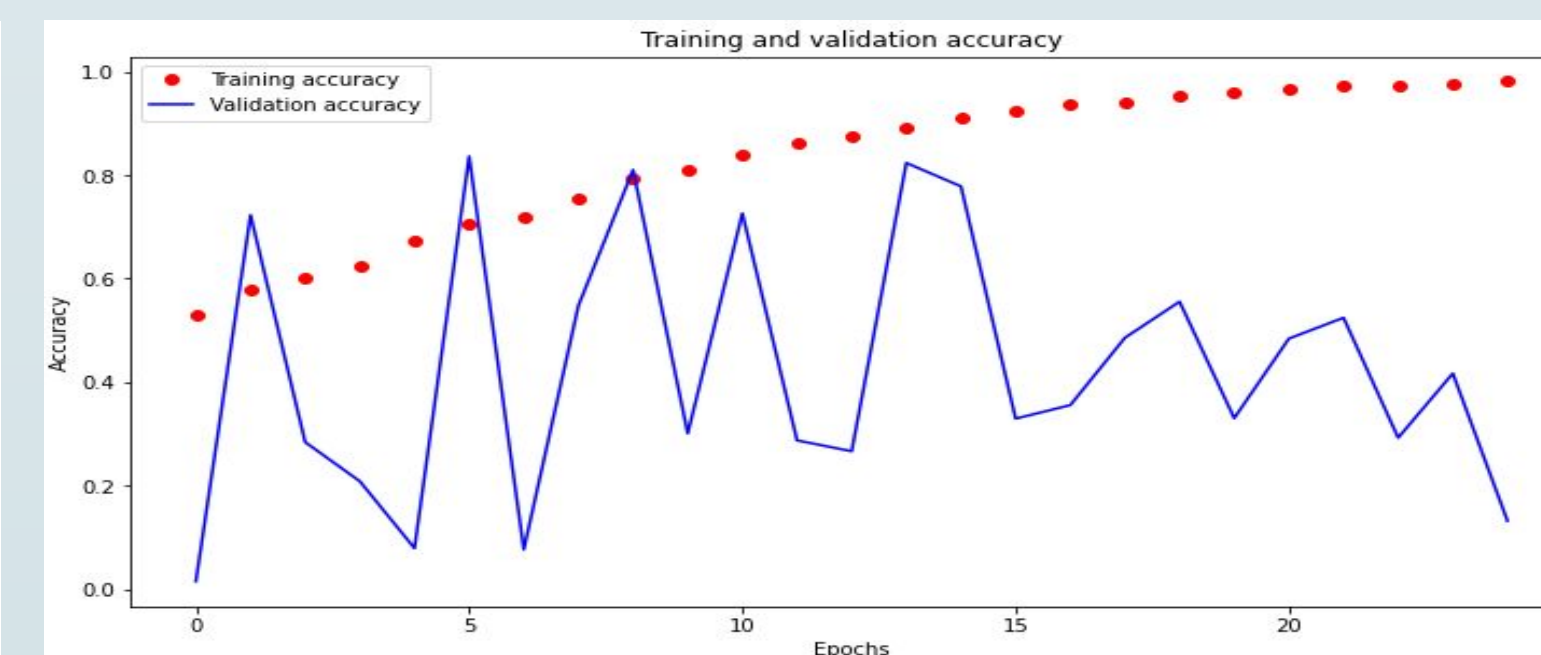
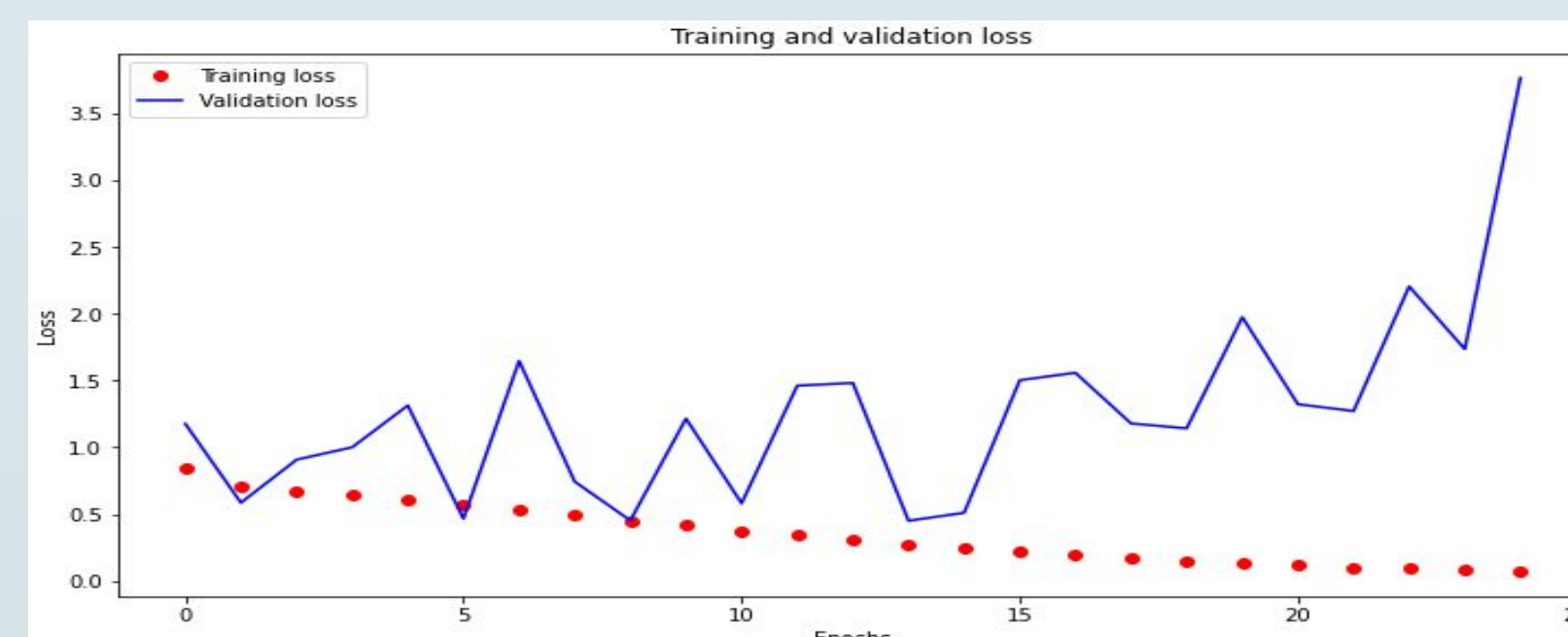


Diagram shows the Schematic of our Data workflow and Network Architecture.  
\*Settled means borrowers that paid back while Past Due means defaulters.

## Methodology

Training a model from scratch requires a very large dataset and takes a lot of time to train. On the other hand, pretrained models have the capacity to converge fast because the weights are already optimized. VGGFace is a series of models developed for face recognition and demonstrated on benchmark computer vision datasets, hence, it is relevant to our task. We have trained our selfies using 3 different CNN models(i.e **our custom CNN model of 4 Conv2d layers, VGG16 and VGGFace+Resnet50**) and the analysis of their performance is shown in the Result section.



Training/Validation Accuracy and Loss of VGG16 Pretrained model.

## Result

Models	Training Accuracy	Validation Accuracy	TP of Settled	TP of Past Due
CNN Model	0.55	0.30	0%	44%
VGG16	0.81	0.34	31%	10%
VGGFace + Resnet50	0.98	0.43	60%	10%

Table shows a comparison of the models on Test set.  
\*TP means True Positive

The two pretrained model were used as feature extractors after which we stacked new layers on top. We observed the following:

- The custom CNN model predicted all 3130 test data as defaulters out of which 44% were correct.
- The VGG16 pretrained model was able to correctly predict 10% of the defaulters.
- In predicting defaulters, the VGGFace model had a matching result of 10%. However, we got a 29% increase for customers that eventually paid back.

## Conclusion

Our pretrained model shows promising results, but performed deplorably on exposure to new selfie data. More so, after several failed attempts on different experiments to yield positive result by predicting borrowers loan repayment status from selfies, we concluded some improvements will be needed. Hence, our future works will entail augmenting the selfies with transactional datasets.