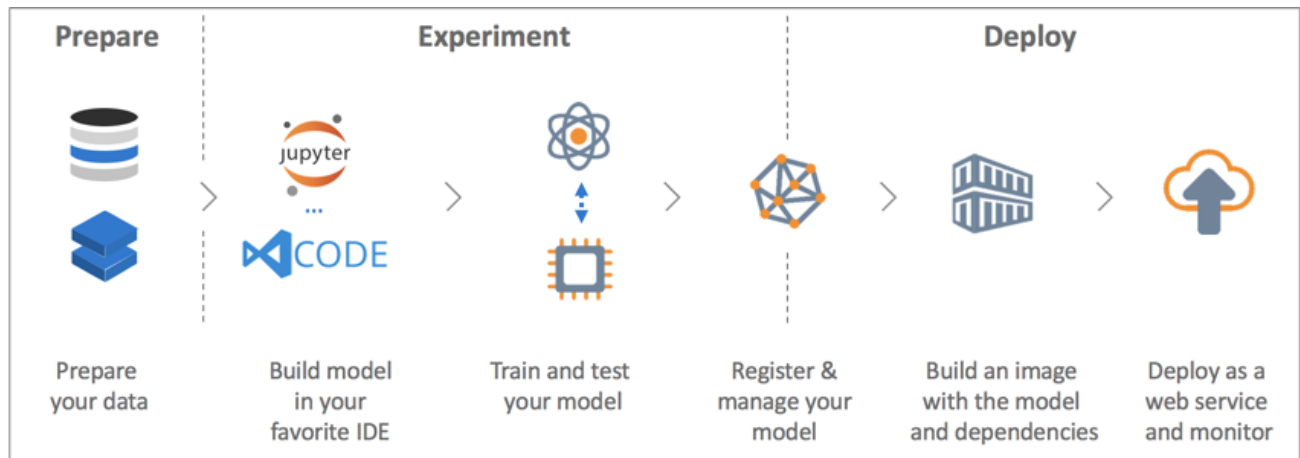# MODELLING

Part 2

*Group Members*
Shrey Raval - 100730265
Jigar Patel - 100730261
Nikhil Singh - 100766644
Tarunjot Singh - 100766653

## Model Architecture and Software Pipeline



Our Software Pipeline has 3 segments as described in the above image. First segment includes the data preparation in which data preprocessing. Preprocessing have several steps like resizing the images, reducing dimension and convert them to vector form. The second phase model is built in Jupyter notebook. Some of the models are VGG16, LeNet-5, CNN. After the model is build, next step is to train and test the model. This is the iterative processing as we need to change Hyper Parameters and features to get the best result. The last phase has further 3 steps. First step is to register the model which means to save and manage the model for further deployment. The last 2 steps are related to model deployment.

## Dataset assumptions, limitations, and constraints

We have used original handwritten MNIST image dataset for our problem. We assume the data set is more than enough of our model to train and predict. The dataset is already cleaned but requires some preprocessing before fitting it for model training. There is no limitation or constraints for this dataset. We are using all the features present in this dataset.

## Scorecard

During the initial stage of the project after applying the Exploratory Data Analysis we came to know that while working with images the most essential part is the type of model that is selected. So, in our case we were able to derive that CNN and Le-net 5 were the two best models that were suitable with our problem any-how we are still checking our compatibility with VGG-16 and Alex-Net for better results.

| Model | Score |
|-------|-------|
| CNN | |
| LeNet 5 | |
| VGG-16 | |
| AlexNet | |

## Selection criteria for model creation and evaluation

For this project we have used CNN and LeNet-5 models as good at classification problems. Moreover, they are less complex and consumes low computational power. VGG-16 and Alexnet are also very good models, but they are more complicated and do not match well with our data set. In addition to that VGG-16 and AlexNet has problem of Vanishing Gradient decent and altogether time and space complexity values are very high compared to CNN and LeNet-5.

## References

1. https://www.azuredevopslabs.com/labs/vstsextend/aml/