

Flow of requests:

- After the trained model is deployed on k8s, User uploads an image of a digit
- Image is received by the application deployed on k8s
 - PIL is used to read the image
 - Image is converted to numpy array
 - Image array is reshaped to have the same shape that the trained model is expecting - **[batch_size, height, width, channels]**
 - Trained model is loaded using Keras API(need to use the same method that was used for saving the model, ex. Keras API)
 - Highest probability score from the model predictions is returned in response to user.
- User receives a json object which has the value as the digit identified.