



# MIDAS@IIITD

---

Multimodal Digital Media Analysis Lab

## MIDAS@IIITD Summer Internship/RA Task 2021

### GENERAL INSTRUCTIONS

Failing to follow any of the instructions below will lead to rejection of your submission.

1. Submission would Include a link to the public github repository.
2. Make sure your code is properly documented. We recommend the following:-
  - a. Before each code block have a markdown block/docstring which mentions the following
    - i. What the code block is doing.
    - ii. What is your intuition behind doing this? Why do you think it is useful?
  - b. Keep an experiment log - document everything that worked or failed. This document(preferably a jupyter notebook) should be a snapshot of the process you follow to solve each problem.
3. Github repository should include a requirement and README file which can be used to reproduce your development environment and code.
4. The individual tasks can have their own requirements.
5. We expect you to complete any of one of three given tasks, whichever one you are most confident with. Feel free to submit more than one task.
6. You have 10 days (Midnight, 10th April IST) to submit the solutions. No extensions will be provided.
  - a. Google Form submission link: <https://forms.gle/PwpyxMopKPX1Yiaj7>

7. If you have any general doubts feel free to email at [midas@iiitd.ac.in](mailto:midas@iiitd.ac.in). In the case of any task specific questions email:
  - i. Task1: [hemantya@iiitd.ac.in](mailto:hemantya@iiitd.ac.in)
  - ii. Task2: [mohits@iiitd.ac.in](mailto:mohits@iiitd.ac.in)
  - iii. Task3: [avinashanandyadav@gmail.com](mailto:avinashanandyadav@gmail.com)
8. Use whatever resources you can find and cite them. Lastly, avoid any plagiarism, if found will be banned from working in MIDAS forever.
9. Selected students will work on project in collaboration with Prof PK and Dr. Rajiv Ratn Shah.
10. Kindly keep checking this file for any additional updates and clarifications.

**Task 1:** For any queries email me at [hemantya@iiitd.ac.in](mailto:hemantya@iiitd.ac.in)

In this competition, you're challenged to use the Speech Commands Dataset to build an algorithm that understands simple spoken commands. By improving the recognition accuracy of open-sourced voice interface tools, we can improve product effectiveness and their accessibility. For more details: <https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/overview>

**Evaluation:**

1. Make a detailed report of the experiments you have done and your observations. Which approach is better and why? It can be as detailed as you want it to be. Include results in the form of tables and figures. Please do not write them in a paragraph. Please restrict it to a maximum of 5 pages including references. Use whatever resources you can find and cite them.
2. Code-base of all the experiments. Follow general instructions strictly.
3. In the README.md mention the rank, score and entry.

**Task 2**

1. Use this dataset (<https://www.dropbox.com/s/pan6mutc5xj5kj0/trainPart1.zip>) to train a CNN. Use no other data source or pretrained networks, and explain your design choices

during preprocessing, model building and training. Also, cite the sources you used to borrow techniques. A test set will be provided later to judge the performance of your classifier. Please save your model checkpoints.

2. Next, select only 0-9 training images from the above dataset, and use the pretrained network to train on MNIST dataset. Use the standard MNIST train and test splits (<http://yann.lecun.com/exdb/mnist/>). How does this pretrained network perform in comparison to a randomly initialized network in terms of convergence time, final accuracy and other possible training quality metrics? Do a thorough analysis. Please save your model checkpoints.
3. Finally, take the following dataset (<https://www.dropbox.com/s/otc12z2w7f7xm8z/mnistTask3.zip>), train on this dataset and provide test accuracy on the MNIST test set, using the same test split from part 2. Train using scratch random initialization and using the pretrained network part 1. Do the same analysis as 2 and report what happens this time. Try and do qualitative analysis of what's different in this dataset. Please save your model checkpoints.

### Task 3: NLP

Assignment Details-

Use a given dataset to build a model to predict the category using description. Write code in python. Using Jupyter notebook is encouraged.

1. Show how you would clean and process the data
2. Show how you would visualize this data
3. Show how you would measure the accuracy of the model
4. What ideas do you have to improve the accuracy of the model? What other algorithms would you try?

.

Note: 1) Goal is to predict the product category.

2) Description should be the main feature. Feel free to use other features if it'd improve the model.

3) Include a Readme.pdf file with approach in detail and report the accuracy and what models were used.

Dataset link:

<https://docs.google.com/spreadsheets/d/1pLv0fNE4WHokpJHUIs-FTVnmI9STgog05e658qEON0l/edit?usp=sharing>

## **Clarifications**