



## Akbank Deep Learning Bootcamp: New Generation Project

In this project, you will be expected to develop a Deep Learning project using basic features.

The aim is to provide participants with practical experience in introductory level classification, data analysis, model development and evaluation techniques in the field of Deep Learning.

Project Topic & Dataset: <https://www.kaggle.com/datasets/crowww/a-large-scale-fish-dataset/data>

With this dataset, you will be expected to develop a classification project.

### 1 - Development Environment (Kaggle&GitHub&README.md)

You need to develop your project in the Kaggle environment on a new notebook that you will create from this data set (by clicking the "New Notebook" button).

**Note 1: Since the dataset size is relatively large at 3GB, it is recommended that you do not download it and work on it locally.**

**Note 2: Participants who wish can also develop their projects in the Colab environment (or any other environment you wish), but we will want the project to be on Kaggle in the final. Therefore, it is recommended that you carry out all development processes on Kaggle.**

- \* You are expected to include technical explanations of your project in your notebook in cells in markdown format.

- \* You should mention your project with summary information in the **README.md** file, which will be located in your GitHub repo.

- \* You should include your notebook link on Kaggle at the end of your README.md file. The outputs of your cells will already be visible and we will be able to run your project on Kaggle if necessary, so **you do not need to add your dataset to your repo.**

## 2 – Data Preprocessing

Your dataset consists of photos in .png format located in folders. Therefore, we cannot read it with `pd.read_csv` as we do in ML applications. You need to organize this data first and create a Pandas DataFrame. The following code snippet is an example:

```
label = []
path = []
fish_dir = '/kaggle/input/a-large-scale-fish-dataset/Fish_Dataset/Fish_Dataset'
for dir_name, _, filenames in os.walk(fish_dir):
    for filename in filenames:
        if os.path.splitext(filename)[-1]=='.png':           # If filename contains .png
            if dir_name.split()[-1]!='GT':                  # If directory doesn't contain GT
                label.append(os.path.split(dir_name)[-1])    # Append the directory name to label
                path.append(os.path.join(dir_name,filename))  # Append all the png files to path of that directory

data = pd.DataFrame(columns=['path', 'label'])
data['path']=path
data['label']=label
```

After this step, **your data** variable is ready as your data set. In order to produce the best project, you can enrich the preprocessing stage by examining the data set with **pandas** commands, or by visualizing the photo data in your data set with the **matplotlib** library.

**Not: To get an idea, you can review the notebooks in the "Code" tab on the dataset's Kaggle page.**

After this step, you are expected to complete this section by separating your data into train and test and preparing it for training.

## 3 – Model Training

You will create a model using ANN architecture. You can get an idea by watching the video in the relevant title on the Education platform and examining the notebook. It will be discussed in more detail in the education publication.

You can include details in your model that will improve your model performance, such as your layers, activation functions, dropout layer, etc.

## 4 – Model Evaluation

In this step, you should include the success metrics of the model. You can use measurements such as accuracy calculation and loss function.

For a better project:

**\* You can plot the graph of your loss function between iterations, create a confusion matrix and classification report.**

## 5 - Hyperparameter Optimization

- Number of layers,
- Number of nodes,
- Dropout rate,
- Optimizer

You are expected to perform hyperparameter optimization by playing with parameters such as and create the best performing model possible.

**Note: Model performance is important in determining the best project. You need to make sure that your model is not overfitted and show and explain this in your code and markdown cells.**