Garden Nerd : Flower Recognition Data Science Competition

HackerEarth

Harshit Singhai
E16CSE147

Shriya Garg
E16CSE054

Devansh Pushkarna
E16CSE167

Presentation Agenda

01 Pre-processing of data set

02 Feature engineering

03 First working model

04 Algorithm and hyper-parameter

05 Summary

06 Further improvements

1

Pre-processing

In association with

DATAQUEST

Pre-processing

- **Image Scaling**: Resized all the image from (500,500,3) to (224, 224, 3) using pre-trained VGG19 and Inception-v4 models
- **Data augmentation:** Basic augmentation like cropping, padding and horizontal flipping were used to increase the amount of data available for better training of models in each category of flowers.
- **Normalizing image inputs:** All the input images were rescaled to 1./255 to make sure all images have similar data distribution and make convergence faster while training network, hence improving time and complexity of the neural network.

2

Feature Engineering

In association with

DATAQUEST

Feature Engineering

Neural Network

- Neural network shifts the burden of feature engineering
- It uses Convolutional neural network for this process.

Deep Learning

- Deep learning has yielded amazing results by learning features and hierarchies on its own.
- The network extracts features during training.

3

Included and excluded features

In association with

DATAQUEST

Included and Excluded features

Zero, zilch, zip, nada, nothing.

- We had no features in our original dataset.
- Hence, we did not have the liberty to add or drop features from the dataset.

4

First working model

In association with

DATAQUEST

First working model

Model from scratch

- Model type used was Sequential() .
- Series of Conv2D and MaxPooling2D layers were used .
- Number of neurons were gradually increased from 32, 64, 128 to 256 for Conv2D layer.
- “relu” activation function was used for all the layers in Conv2D layer
- (3,3) filter matrix was used for all the Conv2D layers
- (2,2) window sized was used for MaxPooling2D layer
- Flatten layers was used to flatten the matrix before the Dense layers

First working model

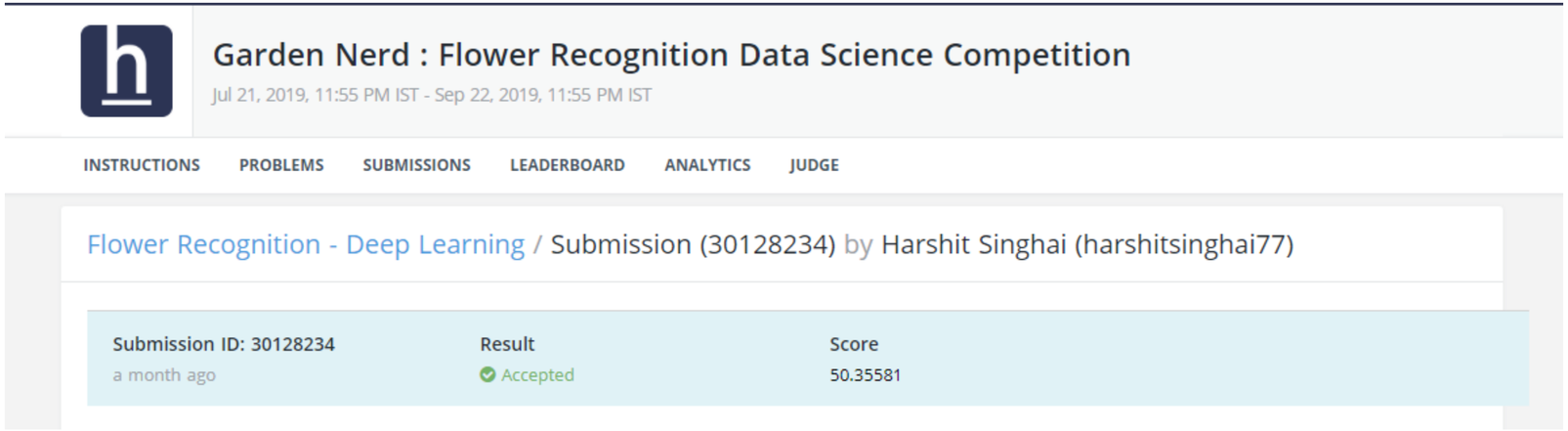
Model from scratch

- Dense layers were responsible for the classification
- 3 Dense layers were used with 512, 256 and 128 neurons respectively.
- “Relu” activation used for the first 2 dense layers
- “Softmax” activation was used for the last Dense layer. Softmax makes the output sum up to 1 so the output can be interpreted as probabilities. The model will then make its prediction based on which option has the highest probability.
- For compilation of the model: ‘binary_crossentropy’ was used as a loss function, RMSprop with lr=1e-4 for optimizer and ‘acc’ for matrix was used

First working model

Model from scratch

- train_generator, steps_per_epoch=100, epochs=30, validation_data=validation_generator, validation_steps=50



5

Selection of Algorithm

In association with

DATAQUEST

Selection of Algorithm

Built from scratch

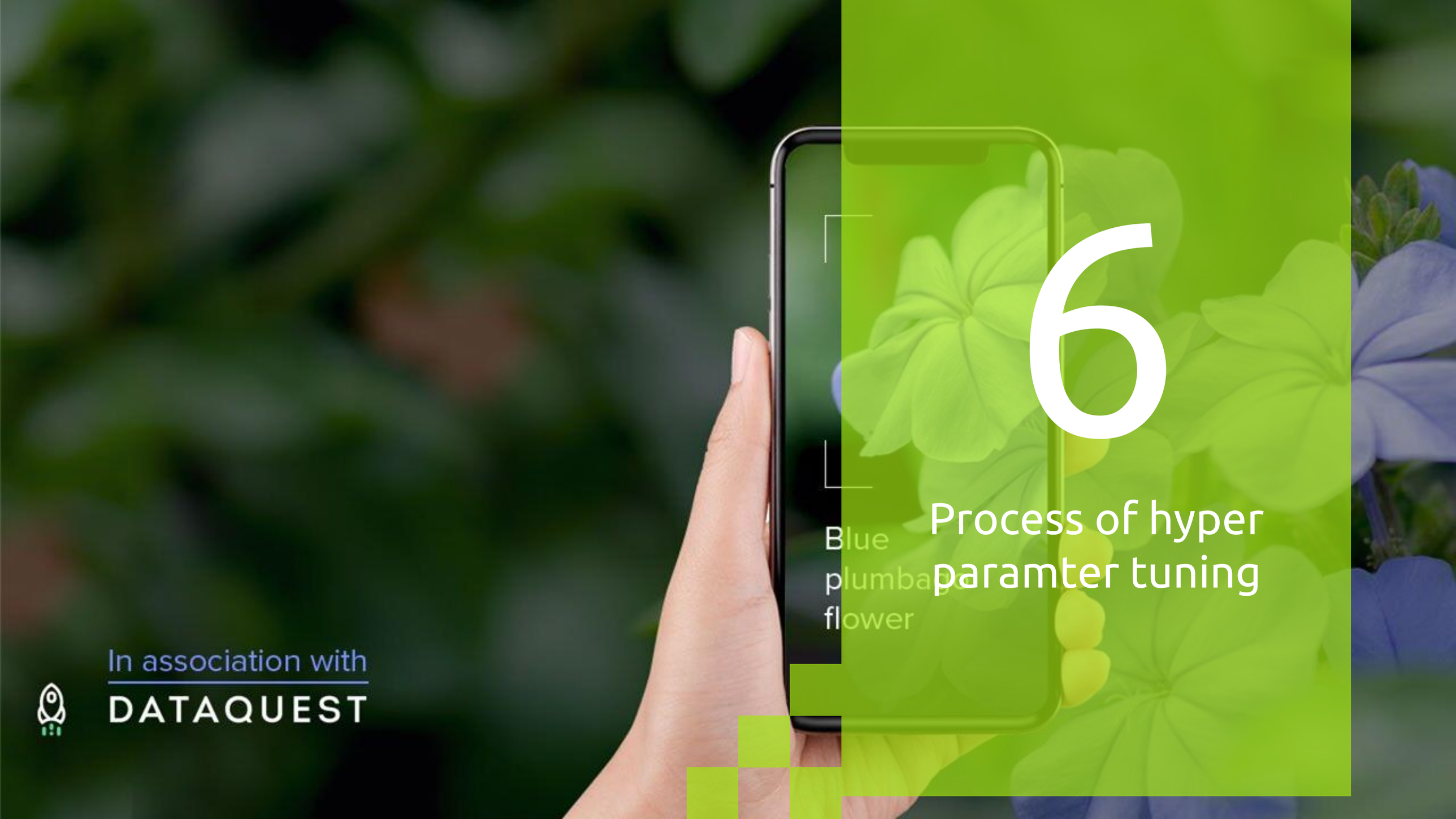
- Initially, data augmentation was not used
- Series of Conv2D and MaxPooling 2D was used to build the model.
- Score was 50%

Feature extraction

- Pre trained models were used for feature extraction
- VGG19
- Resnet 50
- Inception-v3
- Results were not satisfactory

Fine-tuning

- Unfreezed few of the top layers of VGG19 model and fine tuned the classifier as per our needs.
- The newly created classifier was used for further training.



6

Process of hyper paramter tuning

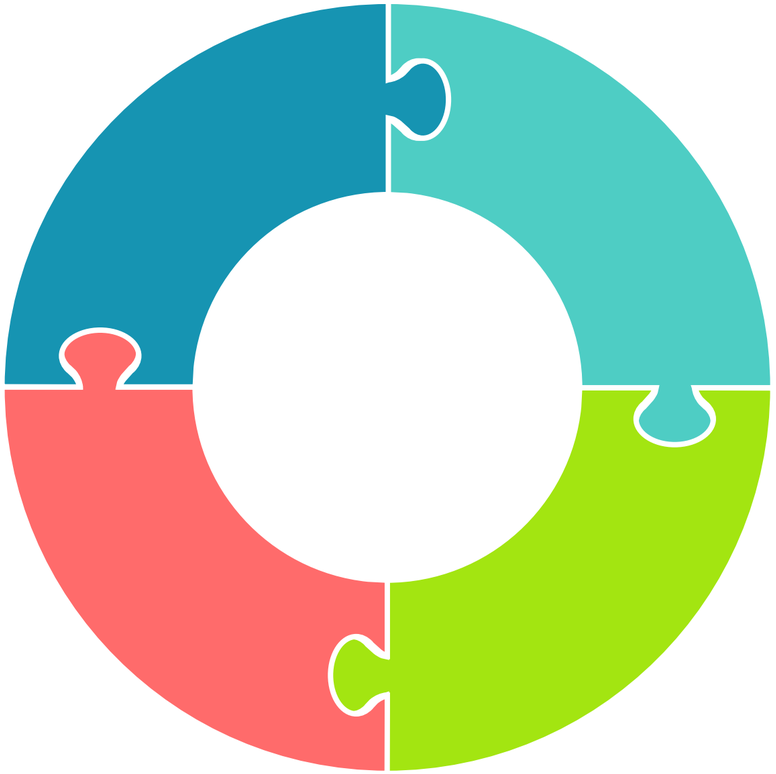
In association with

DATAQUEST

Hyper parameter tuning

Learning rate

Learning rate was crucial in determining how much to update the weight in the algorithm. We tried different learning rate while using RMSprop and Adam optimizer



Number of epochs

Number of epochs are important as they hold a very fine line in underfiting and overfiting the model. More number of epochs can overfit the model which results in model not able to predict the test set accurately.

Batch size

Batch size is imporant for learning process of of convnet. Mostly batch size of 64 and 128 is typically used. We chosed different batchs size for different pre trained model .

Dropout

Dropout is a l2 regularization techniqoe to avoid overfiiting in deep neural networks,. Mostly a default value of 0.5 works good in most of the image classification task.

7

List of hyper parametered values

In association with

DATAQUEST

Tuned hyper-parameter

Batch Size

Batch size with 32 gave the best results.

Steps_per_epochs

Steps_per_epochs around 100 was giving results in reasonable time.

Dropout

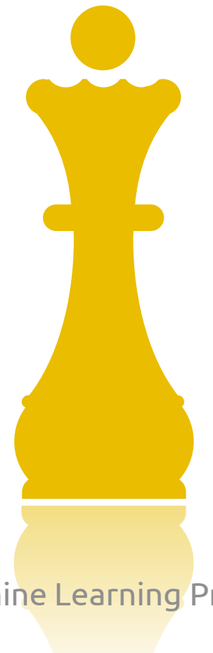
Two Dropout Layers were used with rate of 0.5

Number of epochs

Epoch range between 50-60 were giving good results.

Number of unfreeze layers

For VGG19, last 5 layers were freezed.



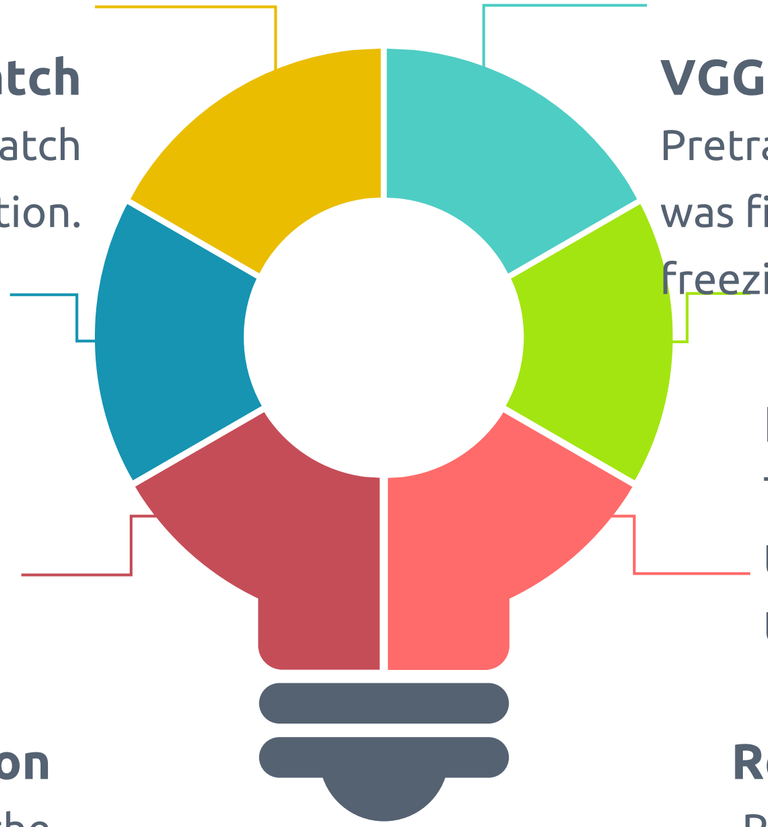
8

How many algorithms tried?

In association with

DATAQUEST

Algorithms Tried



Model From Scratch

The first model was built from scratch without data augmentation.

VGG 19

Pretrained VGG model was used and was fine tuned as per our needs by freezing some layers

Inception v3

Inception was used for feature extraction. The results were quite ok but not great..

Fine tuning

Transfer learning combined with Fine tuning was used on most of the pre-trained model ti improve the results.

Data Augmentation

Basic augmentation of the dataset was used to improve the accuracy

Resnet50

Resnet was used for feature extraction. and further fine tuned.

Attempts (26)

My Submissions

You can choose only 10 submissions for offline evaluation for each problem.

Problem	Result	Score	Detail	Select for Offline Evaluation	
Flower Recognition - Deep Learning	✔	86.06315 (Online score)	view	☐	16 days ago
Flower Recognition - Deep Learning	✔	54.32200 (Online score)	view	☐	19 days ago
Flower Recognition - Deep Learning	✔	54.32200 (Online score)	view	☐	19 days ago
Flower Recognition - Deep Learning	✔	54.32200 (Online score)	view	☐	19 days ago
Flower Recognition - Deep Learning	✔	63.63070 (Online score)	view	☐	19 days ago
Flower Recognition - Deep Learning	✔	89.61446 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	✔	89.61446 (Online score)	view	☐	20 days ago

Attempts (26)

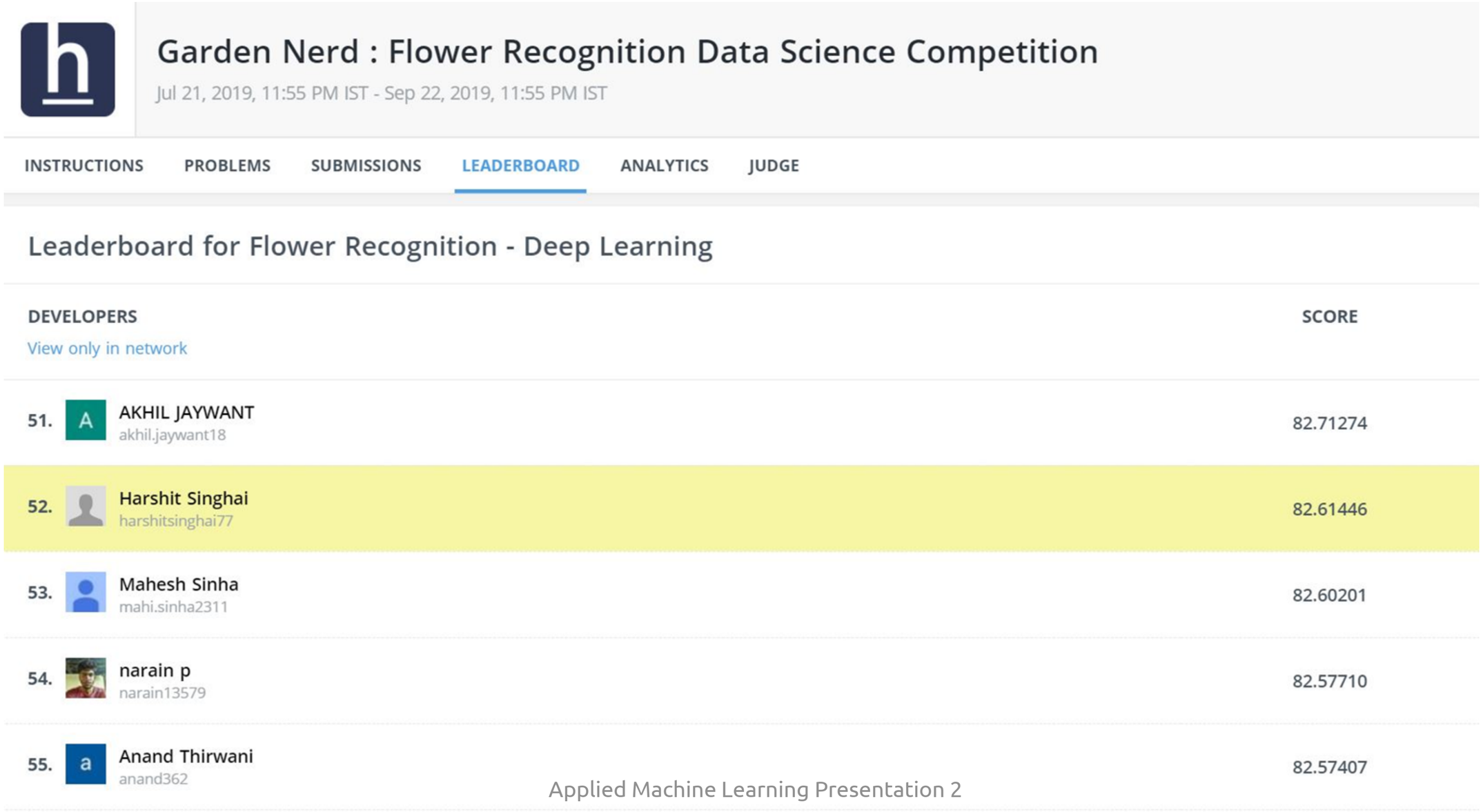
Flower Recognition - Deep Learning	✔	89.61446 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	⚠	0 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	⚠	0 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	✔	9.64897 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	✔	9.28557 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	✔	11.39831 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	✔	10.99820 (Online score)	view	☐	20 days ago
Flower Recognition - Deep Learning	✔	10.04412 (Online score)	view	☐	21 days ago
Flower Recognition - Deep Learning	✔	0.38773 (Online score)	view	☐	28 days ago
Flower Recognition - Deep Learning	✔	2.16426 (Online score)	view	☐	29 days ago

Attempts (26)

Flower Recognition - Deep Learning	✔	2.16426 (Online score)	view	☐	29 days ago
Flower Recognition - Deep Learning	✔	2.16426 (Online score)	view	☐	29 days ago
Flower Recognition - Deep Learning	✔	2.02886 (Online score)	view	☐	29 days ago
Flower Recognition - Deep Learning	✔	0.57920 (Online score)	view	☐	29 days ago
Flower Recognition - Deep Learning	!	0 (Online score)	view	☐	29 days ago
Flower Recognition - Deep Learning	✔	0.08487 (Online score)	view	☐	a month ago
Flower Recognition - Deep Learning	✔	1.05220 (Online score)	view	☐	a month ago
Flower Recognition - Deep Learning	✔	1.05220 (Online score)	view	☐	a month ago
Flower Recognition - Deep Learning	✔	50.35581 (Online score)	view	☐	a month ago

No more submissions to show.

All time best Rank



Garden Nerd : Flower Recognition Data Science Competition

Jul 21, 2019, 11:55 PM IST - Sep 22, 2019, 11:55 PM IST

INSTRUCTIONS PROBLEMS SUBMISSIONS LEADERBOARD ANALYTICS JUDGE

Leaderboard for Flower Recognition - Deep Learning

DEVELOPERS View only in network	SCORE
51. AKHIL JAYWANT akhil.jaywant18	82.71274
52. Harshit Singhai harshitsinghai77	82.61446
53. Mahesh Sinha mahi.sinha2311	82.60201
54. narain p narain13579	82.57710
55. Anand Thirwani anand362	82.57407

9

Summarize the project

In association with

DATAQUEST

Final Leaderboard

Garden Nerd : Flower Recognition Data Science Competition

Jul 21, 2019, 11:55 PM IST - Sep 22, 2019, 11:55 PM IST

INSTRUCTIONS | PROBLEMS | SUBMISSIONS | LEADERBOARD | ANALYTICS | JUDGE

Leaderboard for Flower Recognition - Deep Learning

DEVELOPERS	SCORE
View only in network	
101. Pratik Ahuja (pratik321)	89.61446
102. Jayesh Jain (jjjayesh26)	89.61446
103. Amit Neog (amit3978)	89.61446
104. Harshit Singhai (harshitsinghai77)	89.61446
105. SHRIYA GARG (shriyagarg.123)	89.61446

Advantages

- Helping botanist save time and effort classifying species of flowers.
- Same model architecture can be trained again with more data to classify images other than the initial 102 categories.
- No need for specialist for data collection, hence resources can be saved.

Drawbacks

- Even if the flower belongs to a category beyond the trained 102, it will still be categorised into the given 102.
- Model would not work well for bad quality pictures.
- The flowers will be called with their scientific names, which most general public don’t understand.

10

Approach to further improve the project outcome

In association with

DATAQUEST

Scope of further improvements in project

- Common names of flowers can be added in the dataset, hence making the project more user friendly.
- Time complexity can be further improved to give faster results.
- Trying different combinations.

Thank you!

Any questions ?