

**A  
Report On**

**PROJECT - IV**

**For  
Machine Learning Course  
Fuse.ai Microdegree in AI**



**Submitted By:**  
**Anish Bhusal**  
[bhusal.anish12@gmail.com](mailto:bhusal.anish12@gmail.com)

**Date of Submission: 4th May, 2020**

## **ABSTRACT**

The bee classification dataset consists of two bee species encoded as “1” and “0” where “1” referred to Bumblebee and “0” to Honey bee respectively. Hog features were extracted from images and used as features for training. The dataset was split into train and test sets with 75-25 split rates. Three models: SVM, Decision Tree and Random Forest were chosen for Grid Search with different params. After evaluating the prediction results on the test dataset, SVM algorithm provided better results with 0.791 best score and 80% accuracy.

## 1. DATA EXPLORATION

The given dataset consists of two kinds of bee species: Bumblebee and Honey Bee which are encoded as “1” and “0” respectively.

Total Image Labels	3969
--------------------	------

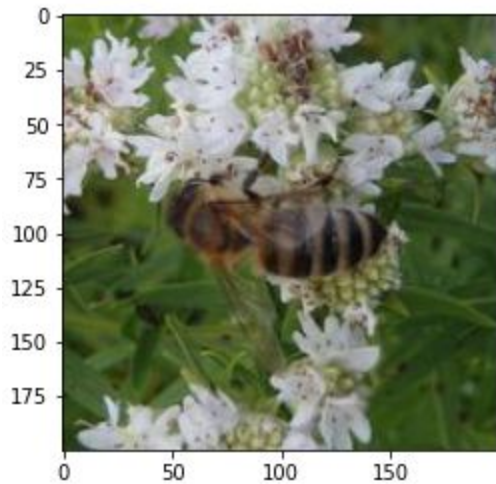


Fig.: Image showing Honey Bee species

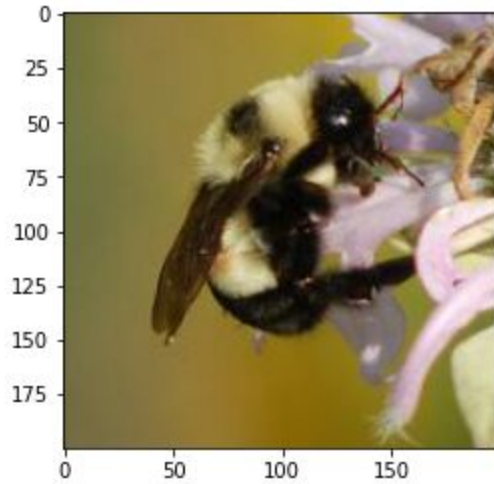


Fig.: Image showing Bumblebee

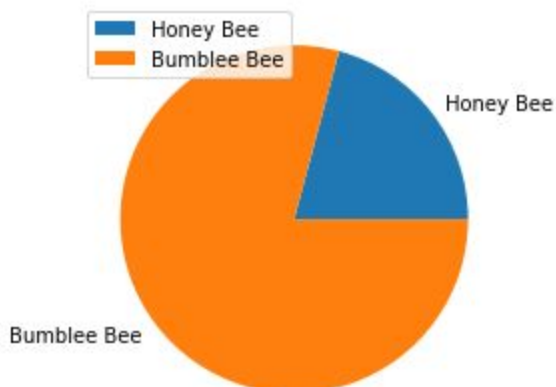


Fig.: Pie-chart showing composition of Honey Bee and Bumblebee in dataset

## 2. DATA PRE-PROCESSING AND FEATURE SELECTION

For each image in the dataset, its hog features were extracted and a dataset X,y was created containing hog features and image labels respectively.

Then the dataset was split in training and test set with following params:

<i>test_size</i>	0.25
<i>random_state</i>	27
<i>stratify</i>	y

## 3. GRID SEARCH

Three algorithms were chosen:

- SVM
- Decision Tree
- Random Forest

With following grid params :

Classifier	Params	Chosen Params
SVM	'kernel':('linear','rbf'), 'C':(2,20)	{ C: 2, Kernel: rbf }
Decision Tree	'max_depth':(9,11,13), 'min_samples_split':(4,6,8, 10)	{ 'max_depth': 9, 'min_samples_split': 6 }
Random Forest	'n_estimators':(5,8,11,13), 'max_depth':(9, 11, 13), 'min_samples_split': (4, 6, 8, 10),	{ 'max_depth': 9, 'min_samples_split': 10, 'n_estimators': 13 }

SVM best score: 0.792

Decision Tree best score: 0.72

Random Forest best score: 0.791

#### 4. MODEL EVALUATION

Three models were evaluated with the help of test set and following classification results were obtained:

	Macro			
Model	Precision	Recall	Accuracy	F1-Score
SVM	0.9	0.51	0.80	0.47
Decision Tree	0.53	0.52	0.72	0.52
Random Forest	0.73	0.51	0.79	0.46

## **5. CONCLUSION**

While evaluating the predictions from three models, SVM provided better classification results with 0.791 best core and 80% accuracy.

Therefore, SVM was chosen to be the best model among three for bee classification.

All the code is available [here](#).