This is a brief overview of the analysis performed on WeRateDog dataset which is based on people comments on different dogs. The data contains features which included tweet ID, timestamp, text, etc. With over 5000 observations.

To further understand the data two addition dataset were used in aiding the analysis, the Image Predictions File which is a table consisting of image prediction whether an image is a dog or not alongside tweet id, image url and number of images used in generating the prediction then the tweet-json file which contain other interesting information about rating.

Among the most highly predicted dog include the golden retriever and Labrador retriever has indicated in the figure below.

```
df["p1"].value_counts(ascending=False).head(10)
```

```
golden_retriever      150
Labrador_retriever    100
Pembroke               89
Chihuahua              83
pug                    57
chow                   44
Samoyed                43
toy_poodle             39
Pomeranian             38
cocker_spaniel         30
Name: p1, dtype: int64
```

Fig 1. p1 prediction

```
df["p2"].value_counts(ascending=False).head(10)
```

```
Labrador_retriever          104
golden_retriever             92
Cardigan                     73
Chihuahua                    44
Pomeranian                   42
Chesapeake_Bay_retriever     41
French_bulldog               41
toy_poodle                   37
cocker_spaniel               34
miniature_poodle             33
Name: p2, dtype: int64
```

Fig 2. P2 prediction

Of all prediction p1 seems to perform better as it prediction falls which within 0.4 - 0.8 which is followed by p2 which has a prediction clustered around 0.1 – 0.2 with a lot of outliers and then p3 which is clustered around 0 – 0.1 and contains outliers as well.

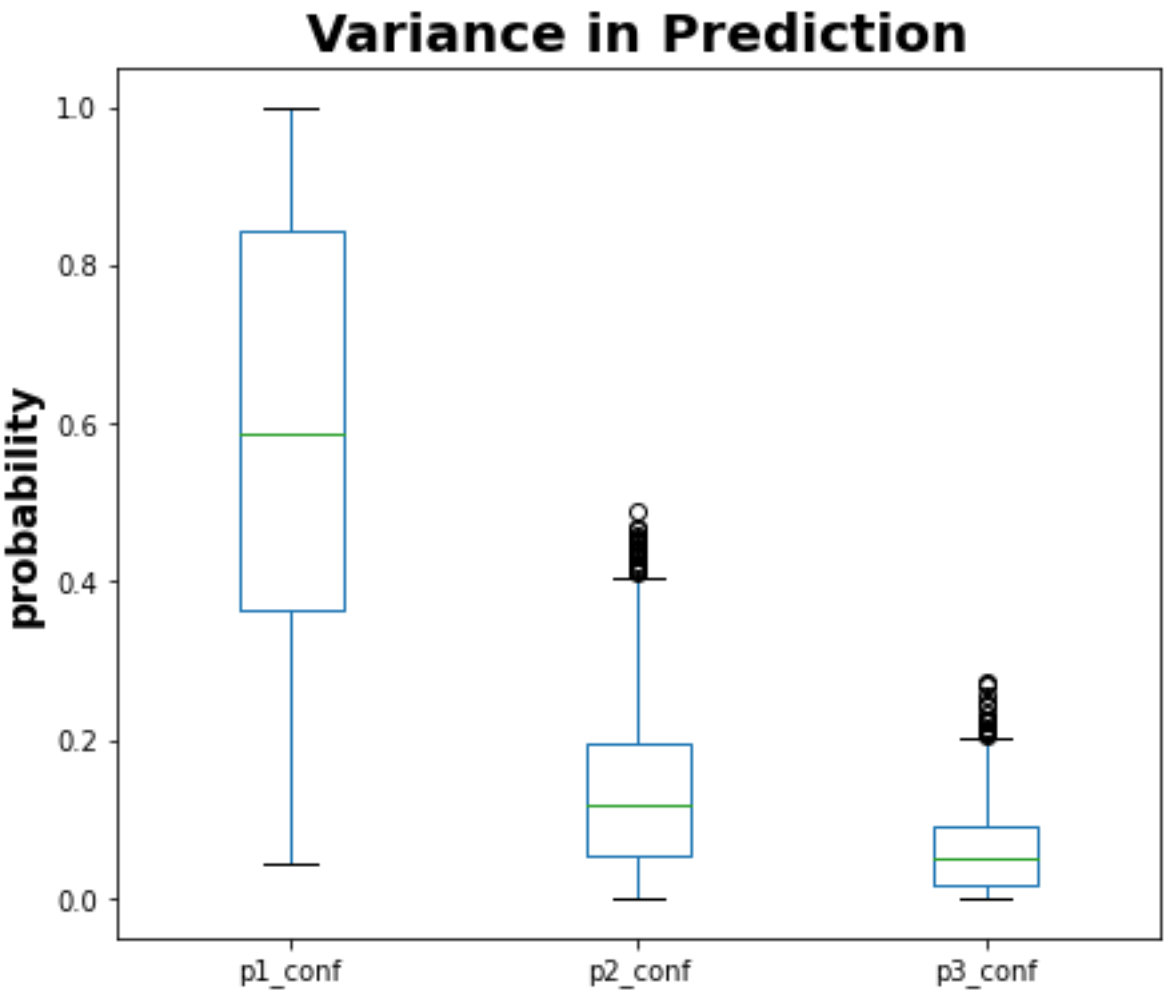

Fig3 histogram probabilities confidence for p1, p2 and p3