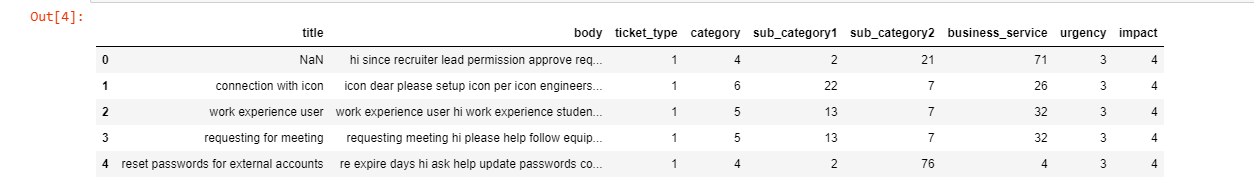
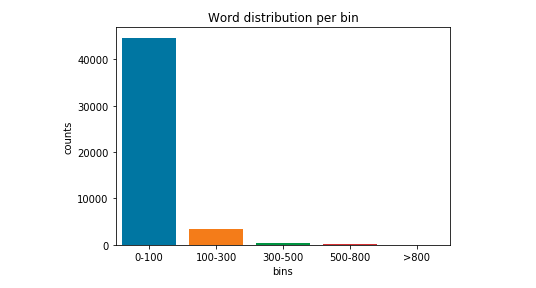
**Behind the Scenes** Analysis for Predicting the Category type, urgency, impact type of Service tickets

Introduction:

A sample of textual service ticket corpus raised by employees of Endava having the details of ticket\_type, category, sub\_categories , urgency and impact were given. A total of more than 48000 documents were recorded.



Binned the number of words in each document to understand what level of description are contained. It has majority of description length within 100 words.



Performed pre-processing starting with simple scikit learn BoW, tfidf followed by lemmatizing and stemming with gensim and then performed topics identification from BoW. With, gensim the word inflections are retained such as universe and university, train and training are considered separate with respect to pos tags. So went with gensim pre processing.

With keras under tensorflow, built with feedforward network the prediction accuracy score on category is 0.81.

When tried the same with LSTM 20 epochs, the accuracy improved and got score of 0.83.

Below are the confusion matrix scores.

precision recall f1-score support

0 0.00 0.00 0.00 1

1 0.00 0.00 0.00 18

2 0.00 0.00 0.00 1

3 0.00 0.00 0.00 47

4 0.87 0.94 0.91 10155

5 0.78 0.71 0.74 2910

6 0.55 0.37 0.44 835

7 0.57 0.42 0.48 273

8 0.90 0.68 0.77 65

9 0.29 0.07 0.11 56

10 0.00 0.00 0.00 1

11 0.62 0.51 0.56 192

12 0.00 0.00 0.00 11

accuracy 0.84 14565

macro avg 0.35 0.28 0.31 14565

weighted avg 0.82 0.84 0.83 14565