**CS4375 Assignment 2**

<https://github.com/carlopizzuto/4375-HW2>

Carlo Pizzuto

cxp200021

# Introduction and Data (5 pt)

In this project, there are two neural networks that need their forward pass functions complete. The first to implement is a Feedforward Neural Network (FFNN), and the second is a Recurrent Neural Network. Both models with be trained for the task of performing a 5-class Sentiment Analysis.

The data being used for this project comes from a set of Yelp reviews. There are three distinct datasets included – Training, Testing, and Validation. Each of the datasets come in a json file, with each object (data point) in all datasets having a ‘text’ variable and ‘star’ variable. The ‘*text’* independent variable is the Yelp review, which will be passed to each Neural Network (NN) to predict the ‘*star’* target variable (ranges from 1 to 5, inclusive).

A graph of blue bars

Description automatically generated with medium confidence

The Train dataset, which will be used to train both Neural Networks, consists of 16,000 data samples, each with its ‘text’ and ‘data’ variables.

In this dataset, all 5 possible target variables are represented equally.

This means that each ‘star’ value has an equal number of data samples.

Figure 1.1 – distribution of target variable **‘stars’** on the **train** dataset.

The Test dataset consists of 800 data samples, each with its ‘text’ and ‘data’ variables. In this dataset, only there are only ‘star’ values of 3, 4, and 5. The ‘star’ value 3 is the least A graph with blue squares

Description automatically generated with medium confidencepresent, with 20% of values having a 3 ‘star’ value. 4 and 5 on the other hand each have 40%.

Figure 1.2 – distribution of target variable **‘stars’** on the **test** dataset.

The Validation dataset consists of 800 data samples, each with its ‘text’ and ‘data’ variables. In this dataset, only there are only ‘star’ values of 1, 2, and 3. The ‘star’ value 3 is the least present, with 20% of values having a 3 ‘star’ value. 1 and 2 on the other hand each have 40%.

A graph of a bar

Description automatically generated with medium confidence

Figure 1.3 – distribution of target variable **‘stars’** on the **validation** dataset.

# Implementations (45 pt)

## FFNN (20 pt)

## RNN (25 pt)

# Experiments and Results (45 pt)

## Evaluations (15 pt)

## Results (30 pt)

# Analysis (bonus: 10 pt)

# Conclusion and Other (bonus: 5 pt)