

## Mini-Project 2: Yelp Business Rating Prediction using Tensorflow

Due Date: 4 pm, Monday, October 4, 2018

### Team Members:

Name: Zainiya Manjiyani  
Student ID: 219216284

### Task Division:

Learned and corrected the mistakes from mini-project-1. Applied label encoding on stars.

Task 1: Consider this problem as a regression problem. Compare the RMSE of the BEST **Tensorflow regression neural network model** you obtained with that of **regression model** you achieved in the last project.

Task 2: Consider this problem as a classification problem. Compare the accuracy of the BEST **Tensorflow classification neural network model** you obtained with that of **each classification model** you achieved in the last project.

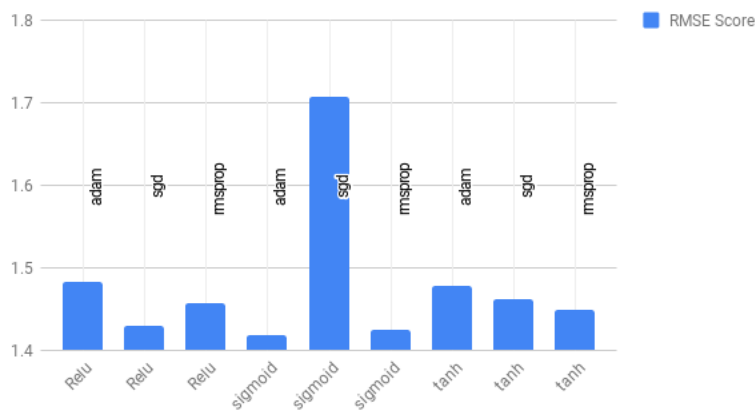
### Project Analysis:

Comparison for Tensorflow regression neural network models:

Activation	optimizer	RMSE Score
Relu	adam	1.482
Relu	sgd	1.429
Relu	rmsprop	1.457
sigmoid	adam	1.418
sigmoid	sgd	1.707
sigmoid	rmsprop	1.424
tanh	adam	1.478
tanh	sgd	1.462
tanh	rmsprop	1.448

Model	RMSE Score
Linear Regression	1.506

Regression



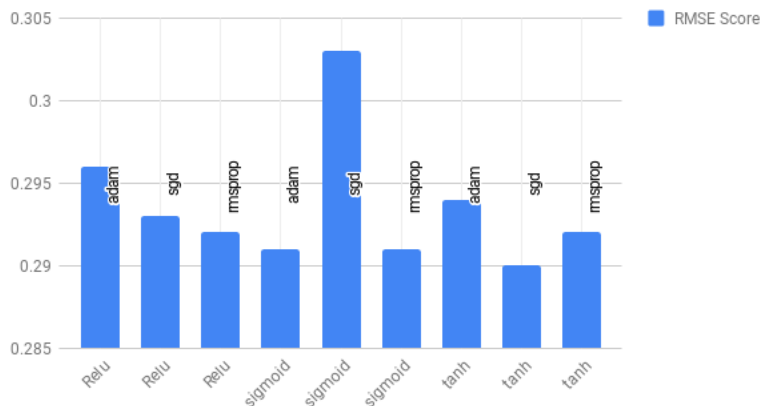
The best result for the regression activation and optimizer combination obtained in sigmoidal with adam, respectively. Linear Regression RMSE Score is better than sigmoidal with sgd.

Comparison for Tensorflow classification neural network models:

Activation	optimizer	RMSE Score
Relu	adam	0.296
Relu	sgd	0.293
Relu	rmsprop	0.292
sigmoid	adam	0.291
sigmoid	sgd	0.303
sigmoid	rmsprop	0.291
tanh	adam	0.294
tanh	sgd	0.290
tanh	rmsprop	0.292

Model	RMSE Score	Accuracy
Logistic Regression	1.667	0.312
Support Vector Machine	1.827	0.272
K Nearest Neighbors	0.412	0.232
Multinomial Naive Bayes	1.667	0.317

Classification



The best result for the regression activation and optimizer combination obtained in tanh with sgd, respectively, having the accuracy of 0.3429.

None of the classification models showed better result than neural network.

And K Nearest Neighbors among all the regression and classification models showed the best performance.

The best result with comparison among all the tensorflow combinations and sklearn models was found in tensorflow combination of tanh with sgd.

### Challenges we faced:

Main difficulty i found in figuring out which type of encoding is needed to require for different algorithms and how to prepare data for neural network model.

I tried to use to\_xy() function to prepare input data in neural network model but i was not able to figure out how encorporate tfidf data in to\_xy() function so I removed it.

## **Learning from the project:**

Data preprocessing mistakes from mini-project-1. In which I took mean of stars and converted into int data type to make models work. But in this project I merged the two data frames based on bussiness\_id and applied label encoding on stars to correct my mistake.

Tensorflow regression and classification for neural network.

Models training effects on the RMSE score.

When and how to apply different types of encoding:

OneHotEncoding and Label Encoding.