Milestone 5 - Results and Interpretation

Team ID: 21

Team Members: Phuong Pham, Jing Lin, Shang-Yun (Maggie) Wu

- Overview
    - We have finished implementing all the models but additional tuning and selection of suitable features are still needed
    - The performance fits our expectation where NN > LR > SVM
    - Tuning doesn't make much difference to the outcome so additional feature engineering or more complex architecture will be tested in attempt to increase accuracy of prediction
- Feature Engineering (completed and being used)
    - Log goal
    - Country
    - Currency
    - Duration weeks
    - Bag of words for description
    - Sentiment analysis on description
- Baseline Models
    - Logistic Regression
        - Regularization: L1, L2
        - Hyperparameters: C, max_iter

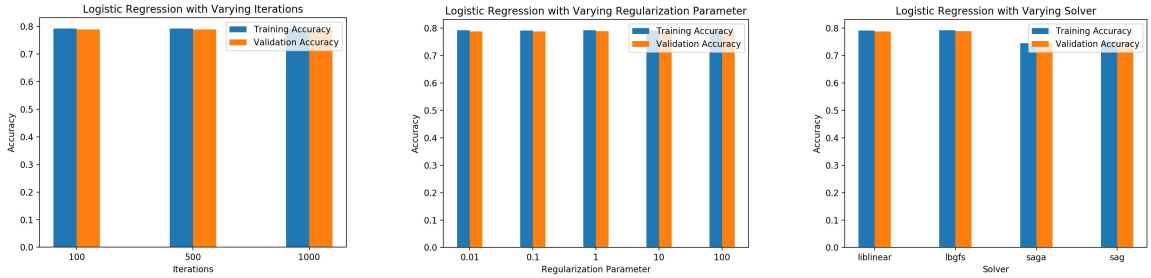

Image 1. Logistic regression performance with respect to varying hyperparameters

- SVM
    - Hyperparameters: lambda

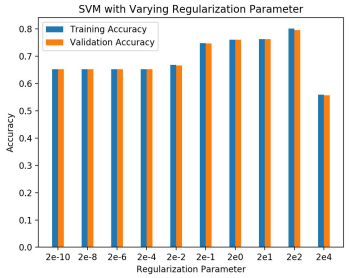

Image 2. SVM performance with respect to varying hyperparameter

- Proposed Models

- Kernelized SVM (Have implemented but no tuning yet)
  - Kernel: Gaussian, polynomial, linear
- Kernelized Logistic Regression (Have implemented but no tuning yet)
  - Kernel: Gaussian, polynomial, linear
- Neural Networks
  - Hyperparameters: hidden layer, neurons
  - Dropout & Batch Normalization
  - Optimizer: SGD, Adam, Adamax, Adagrad
  - Activation functions: relu, sigmoid, tanh

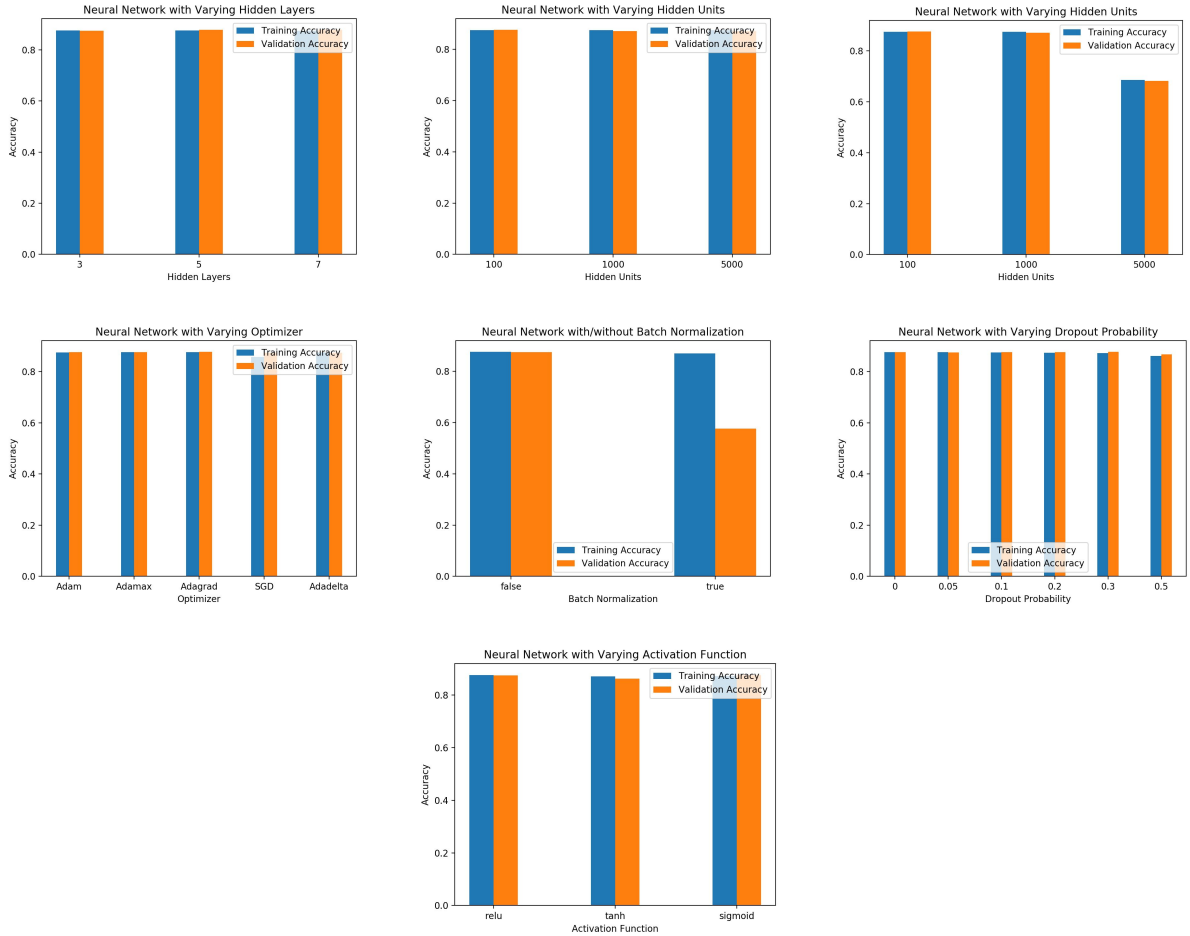

Image 3. Neural network performance with respect to varying hyperparameters

- Comparison (best performance of each model so far)

Model	SVM	Logistic Regression	Neural Network
Training Accuracy	0.801	0.801	0.876
Validation Accuracy	0.796	0.800	0.876

Table 1. Performance comparison across models

- To Do
    - Error analysis
    - Compare and explain the effects of the hyperparameters
    - Performance on test data
- Far-reach Goal (if time permits)
    - Train our own RNN with respect to project description