

sddec18-16: Use machine learning to predict relevant support content based on historical user

Week 6 Report

February 25 - March 3

Team Members

Erin Elsbernd — *Communication Coordinator and Machine Learning Lead*

Ram Luitel — *Project Manager & Software Architect*

Faizul Jasmi — *Testing & AWS Tech Lead*

Khoa Bui — *Database & Web Master #2*

Taizhong Huang — *Testing*

Christian Chiang — *Web Master & AWS Tech Lead*

Summary of Progress this Report

We received labeled data from our client and started processing it into viable training data to use. We also began selecting models to use for the given data and implementing them. We have broken into 2 subteams, where one team will focus on neural networks, and the other on trees and recommendation systems.

Pending Issues

We have some questions about the data since it is spread out over 60 csv files and we are not sure how all of it maps together.

Plans for Upcoming Reporting Period

We will meet with our client to resolve questions about the data and aggregate into a single set of data for everyone to use.

After we finish processing the data we will fit it to our models to get preliminary results.

Individual Contributions

| Team Member | Contribution | Weekly Hours | Total Hours |
|---------------|--|--------------|-------------|
| Erin Elsbernd | Worked on implementing data processing functions such as bag of words and n-grams. Worked on functions to map eventAction sequences to integer sequences. Worked on implementing an LSTM model with text data. | 7 | 43 |
| Ram Luitel | Wrote random forest model to fit out as per out data. The prediction is not correct and not clear what is going on wrong. So I plot the data using Matplotlib to visualize the data to see how the article is behaving in different user activities using plot, pyplot, and pylab. I am trying to use the binary dependent variable, the article helpful or not which represent 1 or 0 the random forest score and | 4 | 34 |

| | the percent helpful as the predicted probability. | | |
|------------------|---|---|----|
| Faizul Jasmi | did research with Pandas on python and tried to use it to parse and clean the data. Wrote the weekly report. Recreated a new timeline for the project plan. | 5 | 32 |
| Khoa Bui | did some research about Pandas. fixed some bugs on the website. wrote design document. | 4 | 32 |
| Taizhong Huang | Worked on the design document and read data. | 4 | 26 |
| Christian Chiang | Did research on Neural Networks. played with some deep learning models. Read other members' code and data. | 4 | 32 |
| | | | |
| | | | |
| | | | |
| | | | |