sddec18-16: Use machine learning to predict relevant support content based on historical user
Week 3 Report
January 27 - February 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 — *Data Base & Web Master #2*
Taizhong Huang — *Testing*
Christian Chiang — *Web Master & AWS Tech Lead*

Summary of Progress this Report

We finally received the data on February 1. It is spread across several CSV files so we are working on combining it together. While waiting for the data we wrote 3 separate programs for classification and feature generation. They included a program compariing different SVM kernels for classifying data, a program to implement a RandomForestClassifier with tuning parameters and charts, and a program using bag of words, and ngrams to create features. These programs were all written in Python using scikit-learn and nltk packages.

Pending Issues

We are still working on how to process the data and get it in a form to use with a classifier.

Plans for Upcoming Reporting Period

Gather the disparate data into one file for use with a classification model. If we can get the data prepared we would like to be able to plot it and see if we can notice any major trends with it. It would also be nice to try to run it through a basic RandomForest classifier and see initial results.

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Erin Elsbernd	Wrote all the models for classification and feature generation. Presented at lightning talk. Scheduled meetings with client.	8	16
Ram Luitel	Created and modified slides for presentation, presented at lightning talk. Worked on team website and updated team page	6	14
Faizul Jasmi	Worked on finding links between data on different files.	5	13
Khoa Bui	Worked on team website and processing data.	5	13

Taizhong Huang	Attended meetings.	4	12
Christian Chiang	Attended meetings, presented lightning talk, updated team site, helped with data processing.	5	13