

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

Week 8 Report

March 18 - March 31

Team MembersErin 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**

This two week we really focus improving our feature extracting process and processing it into viable training data to use. We have finished processing and mapping data. We also did a lot of selecting models to use for the given data and implement them. We have tried different machine learning model to predict the helpful article based on different features. This is the some of the various ML models we tried to use.

1. support vector machines (SVM)
2. k-nearest neighbors algorithm (KNN)
3. Naive Bayes
4. Artificial neural networks (ANN)
5. Random Forest

Pending Issues

Since we have not hit that 70% accuracy prediction requirement for our clients we will continue to work using different machine learning model and different techniques with the different feature to come up with better results. We are going to predict a group of the article instead of a single article and see how the result will different then what we have right now. As of now can predict around 20% accurately.

Plans for Upcoming Reporting Period

We are going to change some of our strategies to implement our predicting models. We are going to predict a group of the article instead of a single article and see how the result will different then what we have right now. Also if we get more data we will redo our features extraction and try some or all of the model that we discussed above to see how the results are different.

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Erin Elsbernd	Finished processing and mapping data. Worked on fitting data to LSTM model and getting more results. Also tried creating Random forest model and fitting data to that. Created metrics to gauge model performance.	8	61

Ram Luitel	I extract the features from the given data and tried using various ML model. Some of them are as follows: support vector machines (SVM), k-nearest neighbors algorithm (KNN), Naive Bayes, Artificial neural networks (ANN), Random Forest. Since it is actually impossible to predict the title ID based entirely on the event transition, I have used all the model to predict the best "id". In fact, the model found some patterns in the features and is able to outperform a random classifier (assigning random labels). The max accuracy was up to around 20% (which is low due to the lack of data and the nature of the state machine itself). Since event transition is a non-deterministic finite state machine.	10	52
Faizul Jasmi	Reading data and consultation from Erin regarding models and results.	4	41
Khoa Bui	Learned how to process data for feature extraction and practice to extract some features from the data. Apply this features using random forest model. Exploring some other machine learning models as well and getting used to playing with data. I am also researching for cleaning data in a data processing pipeline using google cloud.	5	44
Taizhong Huang	Create a model by using KNN, random forest, Decision tree and Naive Bayes.	10	44
Christian Chiang	Research on the best cloud platform to host our Machine learning models. Looking for a platform that is Workiva friendly was the main priority, options was Microsoft Azure, Amazon Web Services, Google Cloud. Little testing on each model to see which one was the fitting	5	42