EE / CprE / SE 492 – Sddec18-16

Use machine learning to predict relevant support content based on historical user interactions.

Week 4 Report
September 24, 2018– October 6, 2018

Team members
Erin Elsbernd: Machine Learning Lead
Ram Luitel: Project Manager and Software Architect
Faizul Jasmi: Communication coordinator
Taizhong Huang: QA Lead
Christian Chiang: Cloud Tech Lead
Khoa Bui: Webmaster and DB lead

Summary of Progress for the Past Two Weeks.

Architecture has been restructured as previous scripts and models consumed .csv files as data input. To speed up the process we are converting all the user data that was given to use in .csv format to DynamoDB tables, this is been done through a lambda function that is designed specifically to convert/translate said files. Scripts and models now have to be changed to consume Dynamo tables instead of the .csv files.

Pending Issues

Our pending issue is same as previous as our model's performance is not where it needs to be in terms of article-group prediction accuracy. We are using the new data that our client has lately provided us. We are expecting after trying couple different model with different features our models will be able to perform at a higher level. We just finished cleaning the new data our next steps is trying to implement different machine learning models and come up with better prediction models.

User data is extremely large, Dynamo tables can take up to 40,000 rows, we are currently at 250,000 rows. If we wanted to store that much data we would need to pay a hefty price and is currently out of budget. We are still looking for a solution to this.

Plans for Upcoming Reporting Period

We are planning to meet our client and advisor next week and update both of them about our progress. And get suggestions and feedback before our second presentation for this semester. We will continue implementing different machine learning models as per our client and advisor suggestion.

Convert the python notebooks to straight up python scripts. Implement the new architecture change to speed up the cleaning process and the data entries. With the new architecture change we need more team coordination.

Individual Contributions

Team Member	Contribution	Biweekly hours	Total hours
Erin Elsbernd	Clean new data and extract features. Try out new data with ML model.	2	12
Ram Luitel	Research on Amazon DynamoDB non-relational database to convert the csv file to .json objects so that we can use in our lambdas to compute in less than 5 minutes.	2	12
Faizul Jasmi	Created a CSV reader using Lambda function to read, process, and store CSV files to DynamoDB	3	14
Taizhong Huang	Learn how to use Dynamo. Tried to make a deep learning model with new data, uncompleted.	4	11

Christian Chiang	Restructure the architecture of the data implementation from previous scripts and models consumed .csv files as data input.	4	15
Khoa Bui	Explore Amazon DynamoDB. Try to learn how to use it for our project proposed.	2	10