**ADS-509 – Team 1 Project Status Update Form**

Fill out this form and submit it by the end of Module 4 in Blackboard.

Team Number: 1

Team Leader/Representative: Dingyi Duan

Full names of team members:

1. Chow, Eva

2. Duan, Dingyi

3. Tan, Abby

Title of your Applied Text Mining project: Tweets Classification on Cryptocurrencies

Short description of your project and objectives:

We are to use Twitter API “Tweepy'' to scrape tweets off the official page of some popular cryptocurrencies and build a supervised machine learning model to classify for corresponding labels (crypto type); Then we are to build a topic model using various text mining techniques to evaluate on how well the topics identified for our tweets lined up with the labels (cryptos types). The designated end goal for deployment is to use “flask” to upload our model(s) to the website (local host) as an end product to predict crypto types based on keyword(s) from user input.

Name of your selected dataset: Approximately 1000 tweets (with retweets) per cryptocurrency, for a total of five cryptocurrencies.

Description of your selected dataset (data source, number of variables, size of dataset, etc.):

We are pulling approximately 1000 tweets per cryptocurrency for a total of five crypto types from twitter API. Using “client.get\_users\_tweets” to perform a maximum of 15 pulls with 100 tweets per pull. Since some cryptocurrencies have less tweets than others, we ended up getting:

For Bitcoin: 1498 tweets.

For Ethereum: 1497 tweets.

For Cardano, 1496 tweets.

For Dogecoin: 957 tweets.

For Shiba Inu: 770 tweets, for a total of 6218 tweets with two columns: “crypto\_type” for crypto types and “tweets” for each tweet we pulled for each crypto.

For this project, use GitHub as a code hosting platform for version control and collaboration. Provide the link here: <https://github.com/dingyiduan7/ADS-509_Final_Project_Team1>

How many times have your members met in the last two weeks? 2

What was the agreed upon method of communication? Are you using any teamwork project management software, such as [Trello](https://trello.com/en-US) or [Asana](https://asana.com/)?

Slack for daily communication and Zooms meeting for weekly updates; Google docs for documentary collaboration and GitHub for code pulling and pushing.

Comments/Roadblocks:

1. Scheduling: All team members are full time working professionals with ongoing personal fairs, the complexity of the project may be affected.
2. Model Performance: The labels we set for our data (tweets) are based on the official pages (crypto types) we pulled from, and since they are all mainstream cryptos, the topics may not be as distinct as we want, which may lead to poor performance of our models.