Steve Setteducati Jr.

DS 4900 Project Pre-Proposal

1. Uses of Spark
   1. Streaming data
      1. Social Media data
      2. Clickstream data
      3. Internet of Things data
      4. Sensors
      5. Web logs
   2. Realtime Analysis
      1. Adjusting predictions and reports as new data comes in, rather than a corpus of large data sets
         1. This still requires a large data set to train on and predictions are made in real time with Spark Streaming
2. Project Idea
   1. Reddit live threads
      1. Reddit supports the ability to stream live updates to their platform
      2. This includes the ability to manage a thread as well as the ability to retrieve discussions on the thread
   2. Analysis on threads
      1. Analysis could be done on the content of a thread, allowing a user of the project to get insights into the type of message being posted on a thread
      2. One could see tags on what the post entails
      3. A relevance score as to if the post will be useful, prior to waiting for a number of likes to help determine that since it is live
   3. A live dashboard
      1. The end product could be some sort of webhosted dashboard with information on a given subreddit
      2. This tool could be used to help give a user insight into reddit posts and see what is relevant faster than reading every post that comes in
      3. Might be useful for companies whose products have a subreddit and therefore have employees working on community relations through them
3. Proposal
   1. By leveraging spark to handle constant streaming data, and live APIs from reddit, I could create a live dashboard running analysis without latency on different subreddits. This dashboard could be used by companies with their own dedicated consumer relations teams to gain faster insight into what people are saying on their forums and help to maneuver complaints, requests, and nonsensical postings more easily. The dashboard would be powered by live classification and NLP models to provide a user with the analytics they want before having to read any post.