Sabya Sachi A20335721
Rohit Jadhav A20337721

Analysis of User Preferences between Android and iOS

Problem

- We are using twitter data to predict the most popular mobile Operating System among users.
- This would help users make decision about the Mobile phone to choose from.
- Also, this would help first time mobile application developers to target larger audience by building application for most popular OS.

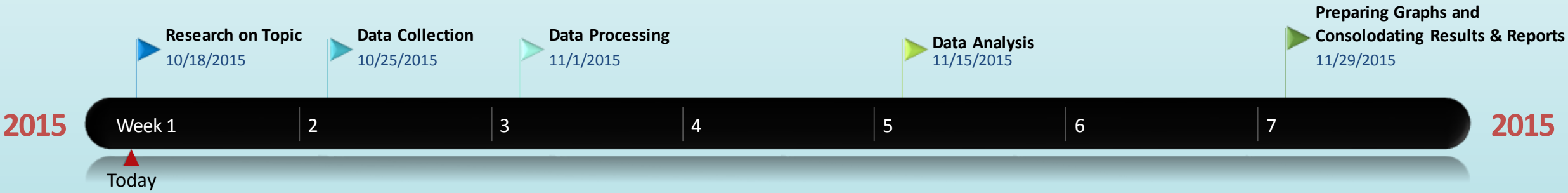
Approach

- Content Retrieval:
  - By using Twitter Streaming API (tweepy)
- Data Preprocessing:
  - Get Clean data and converting to the format as we need
- Data Analysis: NLTK 3.0(Natural Language Toolkit)
  - Popular library for Sentiment Analysis
  - Uses Naïve Bayes Classifier
  - To classify tweet as Negative, Positive or Neutral
- Prediction:
  - Statistics created from previous steps to predict user preferences

Data

- Data Collection :
  - Twitter Streaming API
  - Hashtags to collect relevant data
  - Text field(tweets) for Data processing
- Data Format :
  - According to analysis large amount of noisy, irrelevant and spam tweets in data expected
  - Needs to filter data i.e. removing URLs, duplicates, symbols etc.
  - Required language detection to process only English tweets
  - Manual labelling of data for training dataset

Timeline



Strech –

We will also try to make predictions based on location to make it more relevant.
But this we are keeping as a strech goal – if time permits.