Capstone Proposal

**Predicting Mobile User Demographics of TalkingData**

Adarsh Srinivas

Mentor: Hobson Lane

**Problem to be solved and Motivation:**

Demographics are widely used in marketing to characterize different types of customers. However, in practice, demographic information such as age, gender, and location is usually unavailable due to privacy and other reasons. The task of this challenge is to predict users’ demographics (age and gender) based on their app download, usage, geolocation, and mobile device properties. A postpaid mobile user is required to create an account by providing detailed demographic information (e.g., name, age, gender, etc.). However, a recent report indicates that there is still a large portion of prepaid users (also commonly referred to as pay-as-you-go) who are required to purchase credit in advance of service use. Statistics show that 95% of mobile users in India are prepaid, 80% in Latin America, 70% in China, 65% in Europe, and 33% in the United States. Even in the U.S., the switch to prepaid plans is accelerating during the economic recession from 2008. Prepaid services allow the users to be anonymous—no need to provide any user-specific information. However, building demographic profiles for all customers is critical to mobile service providers. This can help them make better marketing strategies (e.g., identify potential customers and prevent customer churns). Moreover, by using demographic information, service providers can supply users with more personalized services and focus on enhancing the communication experience. Demographic prediction is important for many applications, such as recommendation, personalization and behavior targeting. Doing so will help millions of developers and brand advertisers around the world pursue data-driven marketing efforts which are relevant to their users and catered to their preferences.

**Client:**

**TalkingData** is seeking to leverage behavioral data from more than 70% of the 500 million mobile devices active daily in China to help its clients better understand and interact with their audiences.

Also, one of my friend’s relative is a founder of a similar mobile data service providing company in the Silicon Valley. They are looking to solve some similar problems and using my algorithms developed for TalkingData, I hope to present to him my findings and see if they can be applied to solve some of their business problems. The same algorithms or after certain tweaks to the algorithms, the findings can be used by his company in their marketing efforts, personalizing experience of their users, recommendations and so on.

**Data:**

The data will be obtained from the ongoing Kaggle Competition.

<https://www.kaggle.com/c/talkingdata-mobile-user-demographics/data>

**Approach:**

Initially, I am going to delve deep into the data to get a robust understanding of the different variables. I plan to do some data munging to deal with missing values etc. Once the data is cleaned, I am going to perform some exploratory analysis to recover some interesting insights from the data. Feature engineering will also be applied to discover some new meaningful features that can be used while building the models. Finally, different models will be built using machine learning algorithms. Data visualizations will help in communicating these insights and tell a story.

**Deliverables:**

The capstone project will produce a Presentation (slide deck), a report (highlighting the approach, findings etc.) and the Python Code in the form of a iPython Notebook that will be submitted into my Github repository.