Proposal:

Problem:
I would like to contribute to a current larger research project that seeks to autonomously control UAV Aerial Base Stations and position them optimally in cities to maximize the total communication output for users while minimizing the number of UAV hours. The impetus for this project is newer 5G and 6G networks that require shorter-wavelength frequencies that cannot travel over large distances, requiring more frequent and direct links. In this plan, the UAVs would provide these links to cell towers, increasing internet connectivity for mobile users. I believe that this increased connectivity could pave the way for more connected city infrastructure where buses, trains, and pedestrian traffic management systems can work together to decrease congestion and make travel more efficient.

Proposed Solution:
In this specific project, I want to predict mobile users internet usage from previous data. This means taking a time series of bandwidth and other connectivity measures, and using them to forecast future usage. This prediction would help further align the UAVs with the optimal future locations as each user's usage shifts.

Potential Stakeholders:
- Civilians on the street who desire faster connectivity speeds for navigation, communication, or entertainment.
- Internet of Things devices that require quick connections, like autonomous vehicles, buses, trains, and traffic management systems.

Potential Obstacles:
I am somewhat worried about the data accuracy of my prediction model. I think that the way people use their devices varies a lot, especially between autonomous users and civilians. I think that it will be challenging to find a data set that captures a wide range of users and their data patterns for me to extrapolate from. I am also a little bit concerned about using time series data because I have never done any machine learning with a time series and I am unfamiliar with the different approaches used in that field.

Novelty:
I know that this problem has been considered before in at least one study (https://ieeexplore.ieee.org/document/9417361). However, most of the research focuses on networks of people, instead of predicting individual user's usage. Furthermore, it is important to have a lightweight model for computing network utilization, because any power the drone uses to compute things comes from the power dedicated to flight, thus requiring more drones. My particular application doesn't have to be incredibly accurate to be successful. It is more important for the model to be lightweight and easily computable on the drone. Network usage is fairly consistent, so I don't think this model will have that much of an impact on the broader reinforcement learning model. It is just something more to add to increase the realisticness of the UAV navigation algorithm.

Plan:

Datasets:
I am planning on getting data from this dataset (https://github.com/mariodim/ml_mobile_dataset?tab=readme-ov-file). I may also add additional sources if I need, or if I can find another data resource that more accurately represents the diversity of users in an urban setting. This data is available publicly.

Creating Data:
I may attempt to measure my own internet usage in an urban setting and then include it with the data from the previously mentioned dataset. This depends on if I can find a suitable software to record internet usage.

Data Organization:
My data is a number of different mobile devices, which each have labeled time series data with around 500 data points each. For each second, the data contains information about the bandwidth, round trip time, jitter, de-jittering buffer, and signal-to-noise reduction. I will probably store this in a Pandas Dataframe because it just lends itself to that.

Data Analyzation:
I will first plot each of the components over time to see if there are any obvious correlations over time for any of the features. I will first look for any features that have linear relationships, because those are extremely easy to forecast. If I find any of those common relationships, I will probably just take those features out or extrapolate them with a simple regression model, then use their result for the prediction of the more complex features. However, I expect that much of the data will be very noisy and oscillate quite frequently, so it won't be easily extrapolated by a constant or linear model.

Model Selection:
I have seen an ARIMA model used in a previous paper that addresses the same problem that I am trying to solve. However, I am not sure if this will work because I have multiple features that I am trying to forecast. I may be able to use the ARIMA model for each one, and then cross-validate the results using a linear model that is trained to recognize the relationship of each feature for a specific time point. I also think that there is a library called statsforecast that is supposed to implement ARIMA in an extremely quick way, which will be good for conserving power and computations for the drones.

Model Validation:
To validate my model, I will select a certain prefix of the time series data for training, and then test on the suffix of the time series data. I can repeat this for each of the mobile devices, using a kind of k-fold validation procedure to refine the model.

Similar Models:

As mentioned previously, this paper (https://ieeexplore.ieee.org/document/9417361) will probably be very useful in my work. I also think that this paper which uses the deep learning short-term long-term memory strategy (https://ieeexplore.ieee.org/document/8581000) could help, although I want to stay away from the computational overhead of using deep learning. I will still probably look into this, because I am not that sure how computationally intensive deep learning algorithms are once they are trained, which would be the deployment environment for the UAVs.