Project Background

1. How is this research related to Facebook?
2. Is there any existing research/work done on this topic by anyone from the Facebook team?
3. Will the results of our research be applied / utilised anywhere?

Project Details

1. How can we measure the performance of our model? “True” data for sewage spillage?
2. What data do we need and how do we collect the data?
   1. How do we identify the useful data and access the targeting data?
   2. Which category of satellite images? What features to extract from the images? Frequency and time frames?

Project Expectations

1. What are the technical expectations with regard to the project?
   1. Limited experience in working with time-series analysis and spatio-temporal modeling

Future meeting arrangement

Background

* Dima
  + Data mining,

Water transparency, how much plant in water, plankton

Sally’s Meeting Notes

* Satellite: Daily measure across optical spectrum
  + Transparency
  + How much plant and bloom (plankton)
  + Can probably even take pre-processed data from images with data points, don’t need the raw images
* Idea of project: Whether it is feasible to use satellites to track / monitor sewage pollution
* Ways of approach
  + Doing a correlation study
  + Data:
    - Data from satellite (every location on daily basis)
    - Data on reported sewage accidents
      * Selected locations being monitored, e.g. 400 locations
      * Examples of normal conditions vs pollution accidents
  + Also say something about the way pollution propagates?
* Reason why this can be done
  + Whitepaper published by scientists in response to legislation that water pollution needs to be monitored suggests using satellites
* Practical terms
  + Get satellite data (use tools from previous years)
  + Take data from environment agencies around pollution alerts
  + Start doing some cross analysis - measuring correlation, modelling, formulate it as some supervised learning problem (train model to find pollution, covariates = satellite signals)
* Models to think about
  + CNN,
  + Time series - LSTM, simpler time series models
* Project Outcome
  + Create a method, validate it, with x % of confidence can detect pollution problem
  + Even if just for today, no forecasting
* Comments from Milan
* reasonable spatial resolution, time granularity - daily
* measurement devices at 400 locations
* baselines - naive, from literature?
* null casting (over short periods)
* forecasting by deepmind on weather
* Caveat of using satellite images
  + clouds - satellite images don’t work then
  + extrapolate when there are clouds -
  + correlation of rain with pollution - overwhelm system, so the system will release more polluted water
* Action Plan
  + Google docs: Put together pointers to literature & dataset
  + We can look through the literature
  + List of actions
    - Check dataset
    - Check literature
    - Comment on google docs / email