**DEEP LEARNING PROJECT PLAN SUMMARY**

* **Topic**
* **Research objective / context and motivation**

We will try to classify sounds coming from urban areas. We would like to be able to detect abnormal sounds such as gun shots among more common ones. The idea would be to help the city administration/police detect more quickly dangerous events…

* **Methodology**

The general idea is to transform sound data into images through feature engineer technique such as “Mel” spectrogram (i.e. scale base on pitch) or chromagram (i.e. scale base on pitch categories). This manipulation will allows us to apply a convolutional Neural Networks CNNs and use each feature engineer as a channel to end up with a feature maps. RNN is also a possibility.

* **Dataset**
  + Music Analysis Dataset : <https://github.com/mdeff/fma>
  + Million Song Dataset : <http://millionsongdataset.com/>
  + Open speech : <http://www.openslr.org/12/>
  + VoxCeleb : <http://www.robots.ox.ac.uk/~vgg/data/voxceleb/>
  + Urban Sound Classification : <https://datahack.analyticsvidhya.com/contest/practice-problem-urban-sound-classification/>
  + AudioSet : <https://research.google.com/audioset/?fbclid=IwAR3If9WF29_QwarlvzjwylQVYxxTKNhCAcpA0vanD_hhOe0e8XVfVcyFMYs>
* **Data preprocessing**
  + Need to do some literature review in order to tackle that feature engineering. Could we use a model that already extracts some of these features automatically (AutoInt ?)
* **Estimation method**
  + Leverage some of the feature extraction or feature engineering by using pre-trained CNN model such as the VGG-16 model and then train a fully-connected network.
* **Hyperparameters tuning**
  + Recommendation about the type of hyperparameter we should focus on vs the state-of-art default parameter.
* **Performance measure**
  + Is the accuracy still a good indicator?
* **References**
  + <https://ieeexplore.ieee.org/document/7324337>
* **Questions**

1. What is the best architecture for sound classification? Is the transformation into images something we should consider doing? Are there any other methods known?
2. Do you see any challenges with our project? What would you like to ~~see~~ hear in the final report regarding our project?