SWS3004: Cloud Computing with Big Data - 2018
List of Projects (updated: 5 July 2018)

1. *Cloud-based Recommendation System*

   Objective: Recommendations (marketing) based social media behavior of individual. Example: Obtain insights of an individual's tweets (from Twitter) and recommend movies.

   Tasks: retrieve data from API, filter and process relevant data, define metric(s), analytics, output web application

   Resources
   a. IBM Watson Documentation: https://console.bluemix.net/developer/watson/documentation
   b. Facebook graph API guide: https://developers.facebook.com/docs/graph-api
   c. Twitter API documentation: https://developer.twitter.com/en/docs/basics/getting-started
   d. Twitter API starting guide by W3resource: https://www.w3resource.com/API/twitter-rest-api/
   e. LinkedIn API documentation: https://developer.linkedin.com/docs
   f. Instagram API documentation: https://www.instagram.com/developer/

2. *Personal Internet Footprint*

   Objective: Internet footprint is a social media score that determines the popularity of a person on the Internet. Compute the Internet footprint for an individual. For example, this score can be based on social media presence (counting the number of times social media posts have been shared by other people, the size of the audience that the tweets have reached).

   Tasks: decide on data sources, define metrics(s), retrieve data from API, analytics, output web application

   Resources
   a. Facebook graph API guide: https://developers.facebook.com/docs/graph-api
   b. Twitter API documentation: https://developer.twitter.com/en/docs/basics/getting-started
      Twitter API starting guide by W3resource: https://www.w3resource.com/API/twitter-rest-api/
   c. LinkedIn API documentation: https://developer.linkedin.com/docs
   d. IBM Watson Documentation: https://console.bluemix.net/developer/watson/documentation
   e. Instagram API documentation: https://www.instagram.com/developer/

3. *Cloud-based Automatic Data Visualization*

   Objective: While there is large variety of big data available today, it is not trivial to visualize them for analytics purposes. Given a dataset, this project looks in to automating dataset visualization.

   Tasks: retrieve data, cleaning and preprocessing data, deciding on the type of visualization based on data content, output web application

   Data Sources: data.gov.sg

   Resources
   a. Autovis paper: https://www.cs.uic.edu/~wilkinson/Publications/autovis.pdf
   b. Data.gov.sg data sources: https://data.gov.sg/search
   c. Amazon Open Datasets: https://registry.opendata.aws/

4. *Smart Meal Planner*

Objective: To compose a meal plan that matches nutrition requirements. Example: mix and match different food items to satisfy calorie needs while making sure the mail contains balanced nutrition elements such as proteins and vitamins. You may limit your scope to address special groups such as diabetes patients.

Tasks: Take input from user, retrieve food data and store in database, meal composing algorithm, output web application

Resources
a. USDA Food Composition Database (and API): https://ndb.nal.usda.gov/ndb/
b. FooDB data sources: http://foodb.ca/downloads

5. *Personality Insight Analytics*

Objective: Obtain personality insights for a batch of job applicants and rank them using information from their resumes. Example: application could help a human resources department analyze the personalities of job applicants.

Tasks: data retrieval and organization, integrate IBM Watson service, define metrics to rank applications, output web application

Resources

Data sources:
a. IBM Watson Documentation: https://console.bluemix.net/developer/watson/documentation
b. Demo application: https://personality-insights-demo.ng.bluemix.net/

6. *Combining Cloud Provider Machine Learning Classifiers*

Objective: To improve the accuracy of machine learning classification by combining multiple classifiers. Example: Train models using classifiers from different cloud providers and combine the results to get a more accurate inference result.

Tasks: train models on different platforms, combining classification results-algorithm, output web application

Resources
a. Alexnet - https://www.nvidia.cn/content/tesla/pdf/machine-learning/imagenet-classification-with-deep-convolutional-nn.pdf
b. Nicolas Ding's masters thesis

7. *Performance of Cloud-based Convolution Neural Networks*

Objective: To investigate the impact of pruning Convolutional Neural Networks on accuracy and inferencing time.

Tasks: setup the environment (caffe framework), understand pruning (also how to implement it), acquire data, train the model, record time with different degree of pruning

Resources

a. CNN optimization paper: https://arxiv.org/pdf/1510.00149
b. Alexnet paper: https://www.nvidia.cn/content/tesla/pdf/machine-learning/imagenet-classification-with-deep-convolutional-nn.pdf
c. Image data source: http://www.image-net.org/

8. *Facial Expression Recognition for an Individual (for an Individual)*

Objective: To develop an application to evaluate the facial expression of an individual. Example, tell whether a customer is happy, sad, or, angry after interacting with a customer service officer.

Tasks: filtering/cropping face from the background of an image, train machine learning model for individual faces, get results for individual, output web application

Resources:

a. IBM Visual Recognition documentation: https://console.bluemix.net/catalog/services/visual-recognition
a. IBM Visual Recognition getting started: https://console.bluemix.net/docs/services/visual-recognition/getting-started.html#getting-started-tutorial
   Data sources list1: http://www.face-rec.org/databases/
b. Data sources list2: https://www.behance.net/gallery/10675283/Facial-Expression-Public-Databases
c. Data sources list3: https://www.kairos.com/blog/60-facial-recognition-databases

9. *Facial Expression Recognition (for a Crowd)*

Objective: To develop an application to evaluate the collective expression of a crowd. Example, tell whether the crowd is happy, sad, or, angry.

Tasks: filtering/cropping faces from an image with multiple faces, train machine learning model for individual faces, get results for individual faces cropped from the crowd image, define method of combining expressions of multiple faces, output web application

Resources:

b. IBM Visual Recognition documentation: https://console.bluemix.net/catalog/services/visual-recognition
d. IBM Visual Recognition getting started: https://console.bluemix.net/docs/services/visual-recognition/getting-started.html#getting-started-tutorial
   Data sources list1: http://www.face-rec.org/databases/
e. Data sources list2: https://www.behance.net/gallery/10675283/Facial-Expression-Public-Databases
f. Data sources list3: https://www.kairos.com/blog/60-facial-recognition-databases

10. *Personalized Cycling Route*

Objective: To determine a cycling path that best matches the user's preferences. Eg. Find a circular route that minimizes the altitude gain or a path that mostly runs through national parks and park connectors.

Tasks: take input from user, get map data from APIs, develop algorithm for path planning, output on web application.

Resources
a. Google maps documentation: https://developers.google.com/maps/documentation/
b. Strava API documentation: http://developers.strava.com
c. Singapore NParks data sources: https://data.gov.sg/search?q=NParks

11. *Comfortable Bus Commute*

Objective: To provide with suggestions for bus commuters when to start a journey based on traffic and bus crowd levels. Eg: tell commuter to start journey at 8.30am rather than 8am to reduce travel time and to get a seat on the bus.

Tasks: decide which data use, combining different APIs (gov data, google maps), output web application

Resources
a. Singapore transport API: https://www.mytransport.sg/content/mytransport/home/dataMall.html
b. Google maps documentation: https://developers.google.com/maps/documentation/