UNIL | Université de Lausanne

Project 1 - Big-Scale Analytics

University of Lausanne, Professor: Michalis Vlachos

Deliverable due: Mar 15 23.59pm.

Project description

The goal of this project is to become familiar with cloud based services. We have chosen several services available on cloud providers. You should choose one of these projects (on a first-come-first-served basis) and make a project. We want you to create a comprehensive notebook (where applicable) and video demo of a project using that service and present it in class. This would be beneficial for the whole class as all of you can become familiar with these cloud services thanks to the nice demos presented.

List of topics

Below are the list of topics selected. You should already be familiar with most of the concepts from our previous class. For each topic we have also given you some starting pointers which can help you, but feel free to search for more resources.

1. Sentiment analysis using Google AutoML
   - https://cloud.google.com/natural-language/automl/docs/features

2. Text classification using Google AutoML
   - https://cloud.google.com/natural-language/automl/docs/features

3. Regression using Google AutoML
   - https://cloud.google.com/automl-tables/docs/beginners-guide

4. Machine learning (e.g., regression) with Azure ML
   - ml.azure.com
   - https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machine-learning?context=azure/machine-learning/studio/context/ml-context
   - https://docs.microsoft.com/en-us/learn/browse/?term=machine%20learning

5. Data Cleaning and ML with Azure ML
   - ml.azure.com

6. Chatbots using Microsoft Azure
   - https://docs.microsoft.com/en-us/learn/modules/build-a-faq-chat-bot-with-qna-maker-and-azure-bot-service/

7. Chatbots using Google DialogFlow
   - https://dialogflow.com/

8. Using a NoSQL database (e.g., DynamoDB)
   - https://aws.amazon.com/dynamodb/getting-started/

9. Image classification using Azure
   - https://docs.microsoft.com/en-us/learn/modules/classify-images-with-custom-vision-service/

10. Large-Scale Visualization with Tableau
    - Create a tutorial and an assignment (eg for undergraduate students) based on two datasets. Think of the questions that you want to ask. Create an assignment and the solution. Create a video of the solution as a tutorial.

11. Data Cleaning with Tableau Prep
    - Create a tutorial and an assignment (eg for undergraduate students) based on two datasets. Think of the questions that you want to ask. Create an assignment and the solution. Create a video of the solution as a tutorial.

12. Explainable AI (choose one of the below options)
    - https://xaitutorial2019.github.io/
    - https://github.com/marcotcr/lime
    - Create a notebook to explain what explainable AI means. Use one technique (e.g., LIME given above) to demonstrate the concept with one or two datasets. Don't use only the datasets given by the authors but try other ones as well.

The above will be shown as Project 1 to Project 12 in Moodle. Once you choose one of them it is not available to the other teams anymore.

Deliverables

I. Create a **Python notebook** that explains and demos the concept. Eg if it's about sentiment analysis you should start by describing what sentiment analysis is. Try to make this as comprehensive as possible. Imagine that this is a tutorial to students that want to learn more about the concept. This may include some descriptions of the service your are presenting plus a working example (with code and results) of using the service. It should resemble a blog post which contains introduction, methods or approaches and the code for a working example. (1/3 points)
II. Create a **video** and embed it in your notebook. Imagine you are giving a tutorial on YouTube. (1/3 points)
III. Give a **presentation in class** as if your giving a lecture for approx 20-30 mins. (1/3 points)

Logistics and deadline

1. Select your group in Moodle (up to four people).
2. Select your project in Moodle. Topics are chosen in a "first come first served" basis. So decide quickly.

   The deadline for the submission of your material is **15th of March 2020** at 23h59. After that every week we will dedicate some time for each team to present their project.