Computer Science 4312
Cloud Computing
Fall 2020
Final Project

Date released: Oct. 8, 2020
Submission Deadline: Nov. 22, 2020 (11:30pm)

Goals:

1. Implement a machine-learning (ML) model using a data set in any domain
2. Tune to select the best model
3. Serve the model using FLASK + html
4. Design a Dockerfile to construct a Docker image for the project
5. Deploy the project on an IAAS model
6. Write a report following ACM format

Project:

Search for one data set in any domain by applying some resources mentioned in Lecture 8. Given that dataset, you are required to:

1. Implement an application employing one ML algorithm before tuning to select the best model;
2. Design a website to serve the application using FLASK and html with some features as follows:
   (a) One main webpage to present the application, such as datasets, and ML algorithms, functionalities, etc.;
   (b) One webpage to introduce about the team including information of team members, job allocation, etc.;
   (c) One webpage to serve model's inference;

(d) [**Bonus: 20% points**] One webpage to store/retrieve inferred data to/from Google Cloud SQL.

3. Construct a Dockerfile for deployment preparation;

4. Deploy the project using:

   (a) Google Compute Engine, an IAAS model;

   (b) [**Bonus: 40% points**] Google Kubernetes Engine.

It is expected to submit a report following the ACM two-column format before the deadline. One perfect example of the report's structure is provided at the references section [1].

Deliverables:

- Paper: Two-page (ACM two-column format) including a title, name, affiliation, abstract, introduction, system's description (including an overview of an ML algorithm, IAAS description, etc.), deployment results and conclusions. Make sure to add a list of references (papers, articles, URLs, etc. . .)

- Presentation: few slides that you will use to explain your work including a demo section.

Submission:

- Make sure that your file names follow this naming convention: your paper should be named project_groupID_paper.pdf. Your slides should be named project_groupID_slides.ppt.

- The presentations will be held on the first lecture after the deadline (Nov. 24, 2020) of this Project. However, you are required to submit your files by the deadline (Nov. 22, 2020).

References

[1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 79–86. Association for Computational Linguistics, July 2002.