VIRTUAL MINDS

Applicant Assignments

Data Science & Algorithms Team

Virtual Minds GmbH

The goal of the following tasks is to have a good basis for the discussion in our technical interview. You should not spend hours to prepare a perfect solution, we don't expect a perfectly running system. Take the tasks as a starting point to prepare for the interview.

Outline

We want to discuss with you how to store a machine learning model and access it. You should have received a small python project and a dataset with this pdf.

The file `model_generation.py` generates a model for the given dataset. Don't worry about improving the model since it is only a simple example. We want to focus on the lifecycle of the model, which can be summarized in the following way:

1. The model should be stored after training so that we can retrieve it. For this, we would like to store a unique version with each model. We also want to persist the hyperparameters that were used for training the model and the accuracy of the model. The model should be persisted on a storage like S3. For this, the python script should be extended.

2. The model should be scheduled to be trained once per day. For this we need multiple retries if something goes wrong and also want to define an SLA for the training.

3. We want to monitor the quality of the model. For this we want to visualize the accuracy over time and display the average accuracy of the last week.

4. We want alerting for the following cases:
   - The latest model was generated more than 36 hours ago

- The accuracy of the model exceeds a certain threshold

5. The model should be accessible via an API, so for an input vector it should return the prediction. The API should pull a new version of the model when it is available.

Prepare a Diagram

Prepare a diagram of the components we should use to achieve all of the above goals and how they are connected with each other. For example, which existing software solutions would you propose?

Setup a Component

From the five points listed above, pick one step and run the component. For example, you could choose to set up a software for storing the versioned model and metadata or visualizing the accuracy. Feel free to fake everything you need, **you don't need to set up everything**. E.g. you can fake different models and hyperparameters or accuracy values. The setup should run in a containerized local setup, e.g. with `minikube`.

Last Words

We wish you the best of luck with the exercises and hope that you also have some fun thinking about the tasks! Please make sure to send us all your files (source code, configuration files) together with additional information we need (installation instructions, README, notes, ...). If you have any questions, do not hesitate to contact us.