

Kittyfarm

Mid-Project Review

Highlights

- Fetching the data for all 1.6 million kitties via the Cryptokitties API was no easy task. It took quite a bit of trial and error, but I finally struck the right balance on the number of parameters to call each time to avoid errors.
- The data did not contain the actual gene sequence as I thought it did. This will cause me to pivot from predicting gene outcomes to predicting kitty value directly from cattributes. (Likely gradient boosted trees or random forests approach)

Review Progress

COMPLETED

- Setup EC2, S3, RDS
- Write script to pull data for all 1.6 million kitties from API
- Set up Flask app environment
- Configure data model for kitty features
- Create working sqlite database for test/development

PLANNED

- Create kitty feature table in RDS
- Write script to load features from json files in S3 to RDS
- Build training, validation, and test sets
- Develop CV approach and test potential models (GBM, RF)
- Productionize final model
- Build continuous training/model deployment pipeline
- Build app UI

Demo/Analysis

```
(base) ubuntu@ip-172-31-35-83:~$ aws s3 ls s3://jdc-nu/
#s3.create_bucket PRE test/
2019-05-09 19:51:59 37668925 kitties1.json
2019-05-09 20:06:14 43626632 kitties10.json
2019-05-10 18:55:38 43233151 kitties100.json
2019-05-10 18:57:18 43061628 kitties101.json
2019-05-10 18:59:02 42623274 kitties102.json
2019-05-10 19:00:42 42406986 kitties103.json
2019-05-10 19:02:23 42702842 kitties104.json
2019-05-10 19:04:05 42808283 kitties105.json
2019-05-10 19:05:50 42786982 kitties106.json
2019-05-10 19:07:37 42593445 kitties107.json
2019-05-10 19:09:25 42693524 kitties108.json
2019-05-10 19:11:09 42965474 kitties109.json
2019-05-09 20:07:48 43769253 kitties11.json
2019-05-10 19:12:58 42805385 kitties110.json
2019-05-10 19:14:45 42563296 kitties111.json
2019-05-11 20:15:39 43992343 kitties112.json
2019-05-11 20:42:08 43483507 kitties113.json
2019-05-11 20:44:12 43139155 kitties114.json
2019-05-13 12:35:11 43353827 kitties115.json
2019-05-13 12:37:20 43849143 kitties116.json
2019-05-13 12:54:48 21960264 kitties117.json
2019-05-13 12:55:54 22026377 kitties118.json
2019-05-13 12:57:06 22256670 kitties119.json
2019-05-09 20:09:23 43817737 kitties12.json
2019-05-13 12:58:13 21972827 kitties120.json
```

- 201 json files in S3 bucket (6.36 GB) containing data for 1.6 million cryptokitties

Lessons Learned

- This project has been an awesome opportunity to learn more about how Dapps work on the blockchain. Diving into cryptokitties has helped me garner a better understanding of blockchain tech.
- Building the framework for a Flask app in Python has been very helpful. At AeroPay, the company I am working for, uses a Flask app to facilitate the back-end of our user and merchant apps.
- At an even broader level, the development work in Python has been boosting my Python skills. It is such a versatile language, and I have been needing to build my Python skills to the same level as R.

Recommendations

- Load data from S3 to RDS
- Build training, testing, and validation sets
- Explore GBM and RF models for predicting kitty value
- Productionize final predictive model
- Build continuous training/deployment pipeline
- Finalize app UI