



A survey of cloud-based machine
learning services

6 DECEMBER 2018

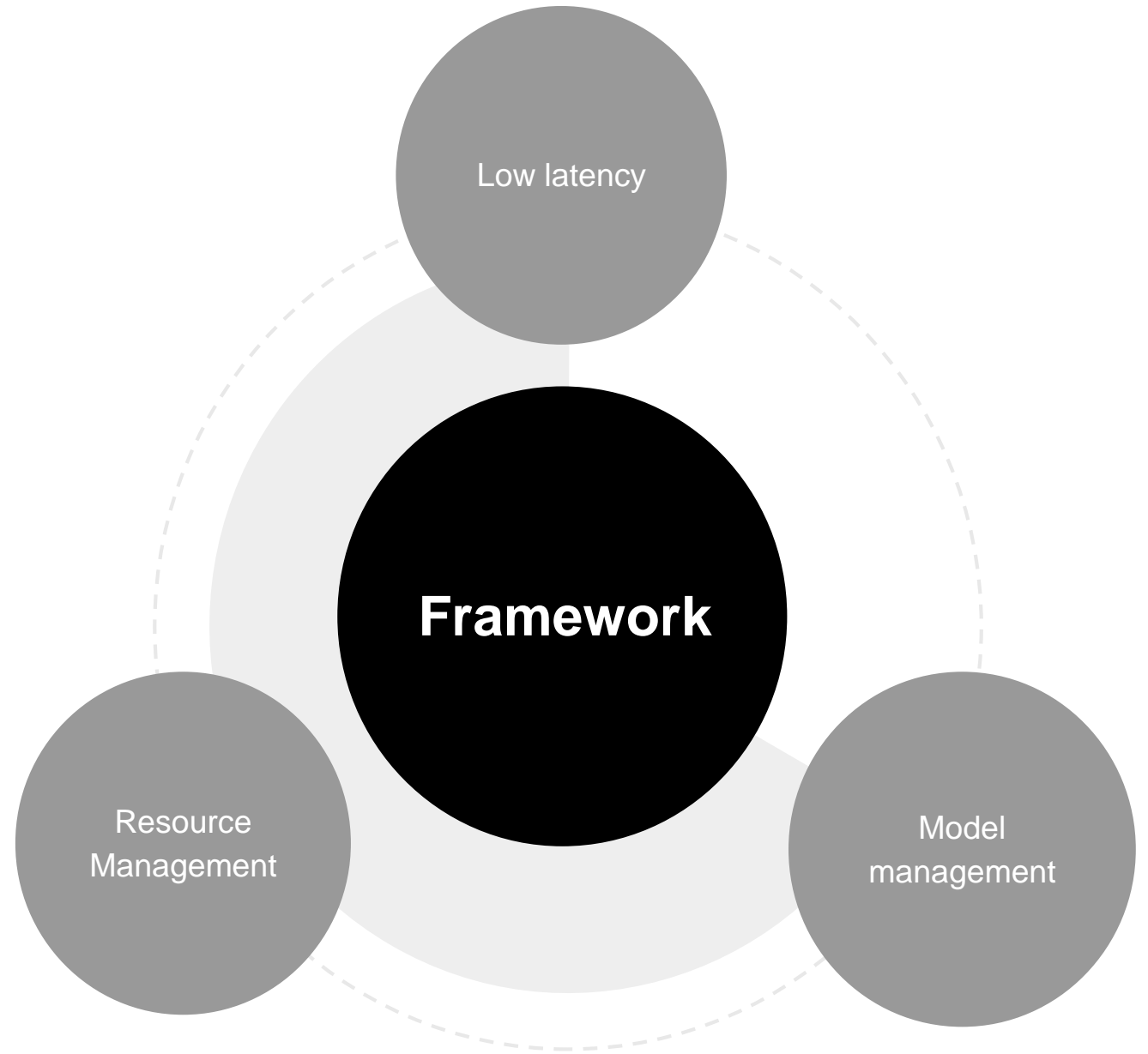


A large white circle is centered on a black background. To its left, there is a series of overlapping circles in various shades of gray, with the number '1' in white at the center of the darkest circle. To the right of the white circle, there are several concentric white circles of decreasing size.

1

Criteria

Framework
needs to have
the following
criteria





2

Cloud Machine Learning Services

Azure Machine Learning Studio

Ease of use

Has a wide range of built-in dataset for developers as well as data scientists to experiment. It also has a large public library models for us to play around with.

Intuitive UI

Azure gives you a very intuitive dashboard for you to play with, there is a experiments section where you can perform data analysis and training model with their built-in dataset.

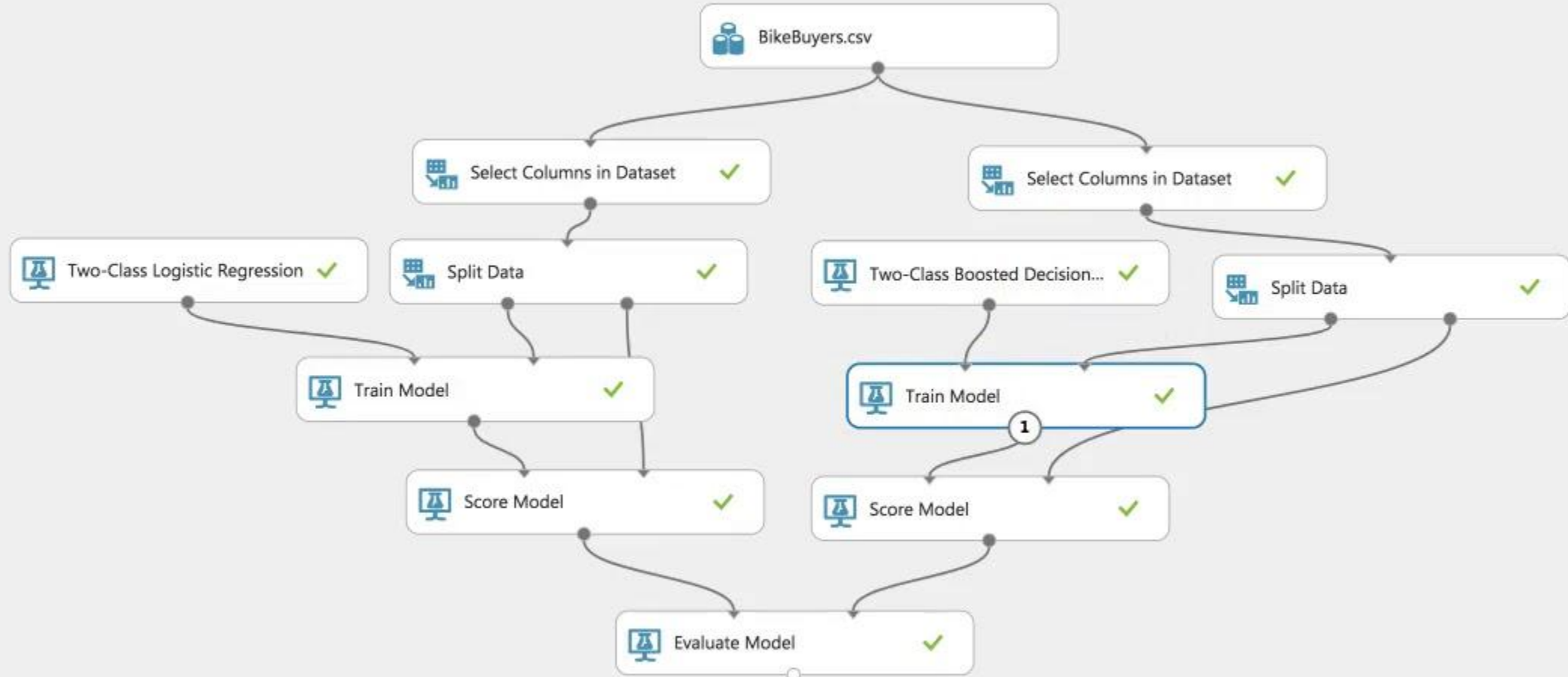
Large built-in models

It has more than 100 methods that address classification (binary+multiclass), anomaly detection, regression, recommendation, and text analysis.



Evaluate Model: Bike Buyers

Datasets, Modules, Trained Models, and Transforms



Azure Machine Learning Service

- Supports environment for building model, experiment with them, and use a broad variety of open source components and frameworks.
- Does not have built-in models, requires custom model engineering.
- Supports popular frameworks such as TensorFlow, scikit-learn and Microsoft Cognitive Tool as well as Docker for deploying model into production.



Google Cloud Datalab

Integrated & Open Source

Cloud Datalab is built on Jupyter, which boasts a thriving ecosystem of modules and a robust knowledge base. Cloud Datalab enables analysis of your data on Google BigQuery, Cloud Machine Learning Engine, Google Compute Engine and Google Cloud Storage

Scalable

Whether you're analyzing megabytes or terabytes, Cloud Datalab is able to query terabytes of data in BigQuery, run local analysis on sampled data and run training jobs on terabytes of data in Cloud Machine Learning seamlessly

Data Management & Visualization

Use Cloud Datalab to gain insight from your data. Interactively explore, transform, analyze, and visualize your data BigQuery, Cloud Storage and Python



Google Cloud ML Engine

- Automatic resource provision feature helps us to focus on model development and deployment without worrying about infrastructure.
- HyperTune feature enables automatically tuning deep learning hyperparameters.
- Support multiple frameworks to train and serve classification, regression, clustering, ...
- Serving model for batch and online learning is very easy.



Amazon SageMaker

- Provides Jupyter notebook for data exploration and analysis without server management hassle.
- Offers built-in models for production purposes.
- Runs on a fully managed elastic compute server.
- Supports most of the frameworks such as MXNet, Keras, TensorFlow, Pytorch, Caffe2, ...



Service Evaluation

- All 3 services support resource management and model management well.
- Model experimentation is easy on all services.
- Serving models is much easier with Google Cloud ML since it does not involve docker knowledge.



VS



VS



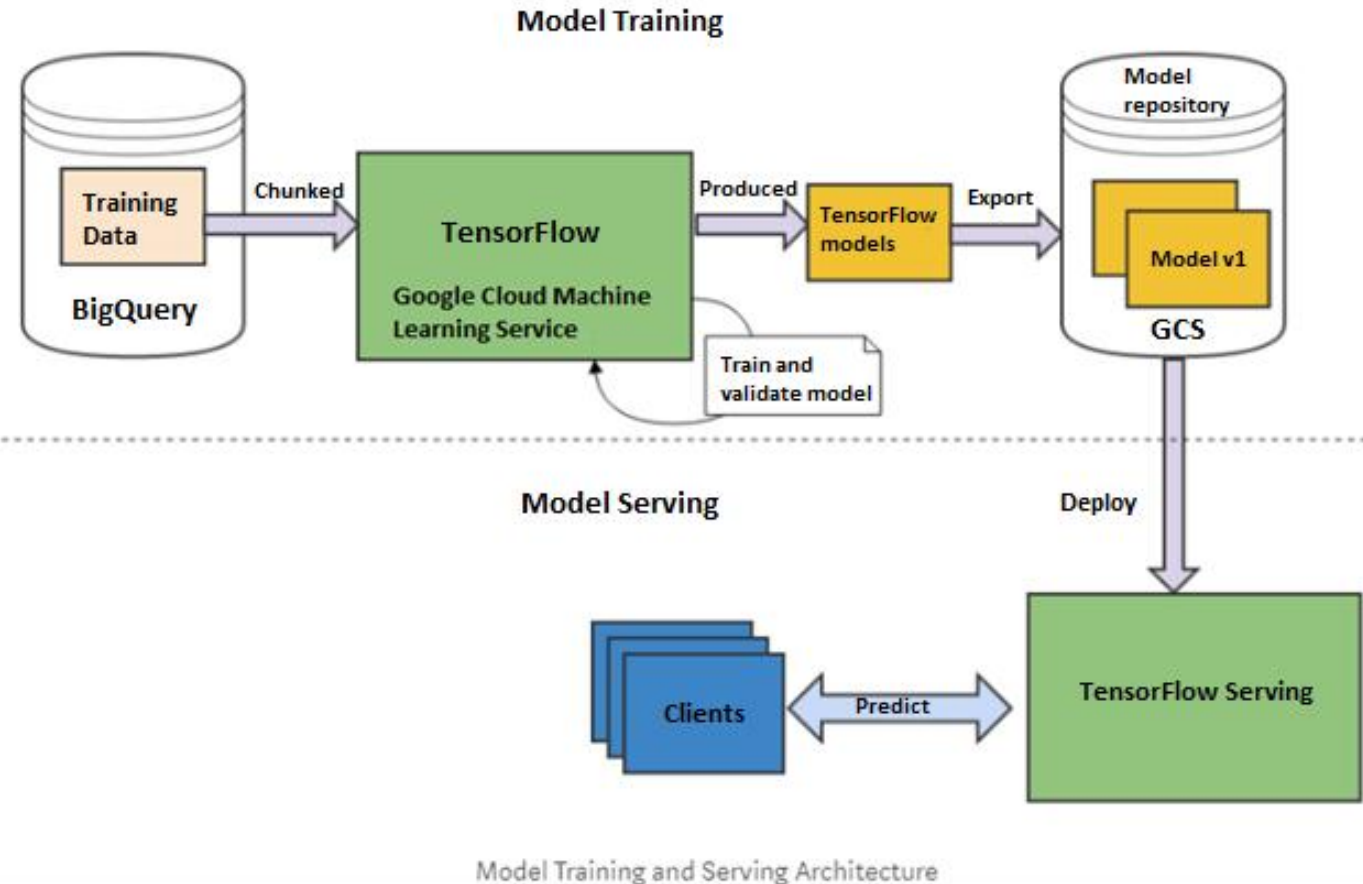
A large white circle is centered on a black background. To its left, there is a series of concentric circles in shades of gray, with the number '3' in white at the center. To its right, there is another series of concentric circles, this time in white, also centered on the black background.

3

Solution

Solution

- Use BigQuery for storing training data
- Use Google Cloud ML Engine for training models
- Use Google Cloud Storage for storing models
- Use TensorFlow serving for prediction





Thanks!

Any questions?

You can find me at

- bao.bui@knorex.com