- Deploy the model.

```
%%bash
MODEL_NAME="superconductor"
MODEL_VERSION="v1"
MODEL_LOCATION=gs://$bucket_name/jobs/
superconductor_181222_040429/4/export/superconductor/1545452450

echo "Deploying model $MODEL_NAME $MODEL_VERSION"
gcloud ai-platform models create ${MODEL_NAME} --regions us-central1
gcloud ai-platform versions create ${MODEL_VERSION} --model
${MODEL_NAME} --origin ${MODEL_LOCATION} --runtime-version ${tf_
version}
```

Batch Prediction

The following code carries out inference on the deployed model:

- Submit a batch prediction job.

```
%%bash
JOB_NAME=superconductor_prediction
MODEL_NAME=superconductor
MODEL_VERSION=v1
TEST_FILE=gs://$bucket_name/preproc_csv/data/eval-00-of-01.csv
OUTPUT_DIR=gs://$bucket_name/jobs/$JOB_NAME/predictions

echo $OUTPUT_DIR

# submit a batched job
gcloud ai-platform jobs submit prediction $JOB_NAME \
        --model $MODEL_NAME \
        --version $MODEL_VERSION \
        --data-format TEXT \
        --region $region \
        --input-paths $TEST_FILE \
              --output-path $OUTPUT_DIR
```

```
# stream job logs
echo "Job logs..."
gcloud ml-engine jobs stream-logs $JOB_NAME
```

```
'Output':
gs://superconductor/jobs/superconductor_prediction/predictions
Job logs...
INFO    2018-12-22 22:04:22 +0000   service     Validating job
                                                requirements...
INFO    2018-12-22 22:04:22 +0000   service     Job creation
                                                request has been
                                                successfully
                                                validated.
INFO    2018-12-22 22:04:22 +0000   service     Job superconductor_
                                                prediction is
                                                queued.
INFO    2018-12-22 22:09:09 +0000   service     Job completed
                                                successfully.
```

- List the contents of the prediction output directory in GCS.

```
%%bash
gsutil ls gs://superconductor/jobs/superconductor_prediction/
predictions/
```

```
'Output':
gs://superconductor/jobs/superconductor_prediction/predictions/
prediction.errors_stats-00000-of-00001
gs://superconductor/jobs/superconductor_prediction/predictions/
prediction.results-00000-of-00002
gs://superconductor/jobs/superconductor_prediction/predictions/
prediction.results-00001-of-00002
```

- Show predicted RMSE outputs.

```
%bash
# read output summary
echo "Job output summary:"
gsutil cat 'gs://superconductor/jobs/superconductor_prediction/
predictions/prediction.results-00000-of-00002'
```

```
'Output':
{"outputs": [0.02159707620739937]}
{"outputs": [0.13300871849060059]}
{"outputs": [0.02054387889802456]}
{"outputs": [0.09370037913322449]}
                ...
{"outputs": [0.41005855798721313]}
{"outputs": [0.39907798171043396]}
{"outputs": [0.4040292799472809]}
{"outputs": [0.43743470311164856]}
```

This chapter provided a walk-through of an end-to-end process to model and deploy a machine learning solution on Google Cloud Platform. The next chapter will introduce the concepts of a microservice architecture. It provides an overview of working with Docker containers and their orchestration with Kubernetes on GCP.

PART VIII

Productionalizing Machine Learning Solutions on GCP