Loan Eligibility Prediction Project using Machine Learning on Google Cloud Platform

Dr. Aygul Zagidullina

March 14, 2024

Introduction

Using Python to build a predictive model on GCP to determine whether an application requesting loan is eligible or not

What will we learn?

- What is loan eligibility prediction, and why is it important?
- What is GCP?
- What are the types of virtual machines available in GCP?
- How to create a virtual machine in GCP?
- How to set the IP address as Static for a VM instance in GCP?
- How to configure Jupyter in the GCP virtual machine?
- How to build Classifier models with Python?

Problem Statement

Business Objective:
The primary source of profit for the banking sector is from loan interests. Banks invest a lot of time and money in analyzing a loan applicant's profile. Going through the profile of every applicant manually is a highly tedious job. This is why automating this process is necessary.

Aim: To build a predictive model to identify if an applicant is capable of repaying the loan or not.

Tech Stack:
Language: Python
Packages: Pandas, Numpy, Scikit-Learn, XGBoost, Imblearn
Cloud: GCP

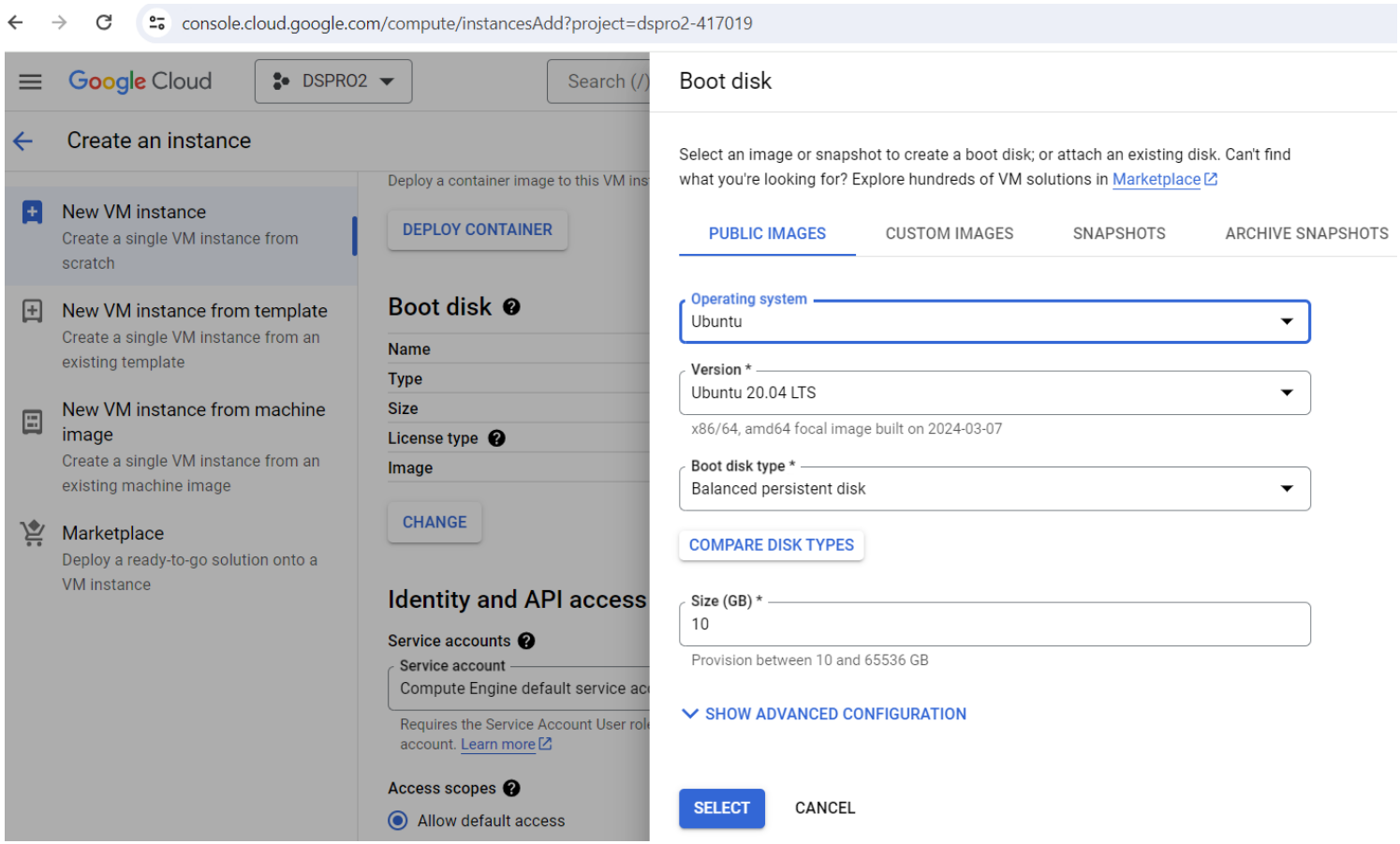
Instance Creation on GCP

3 ways:

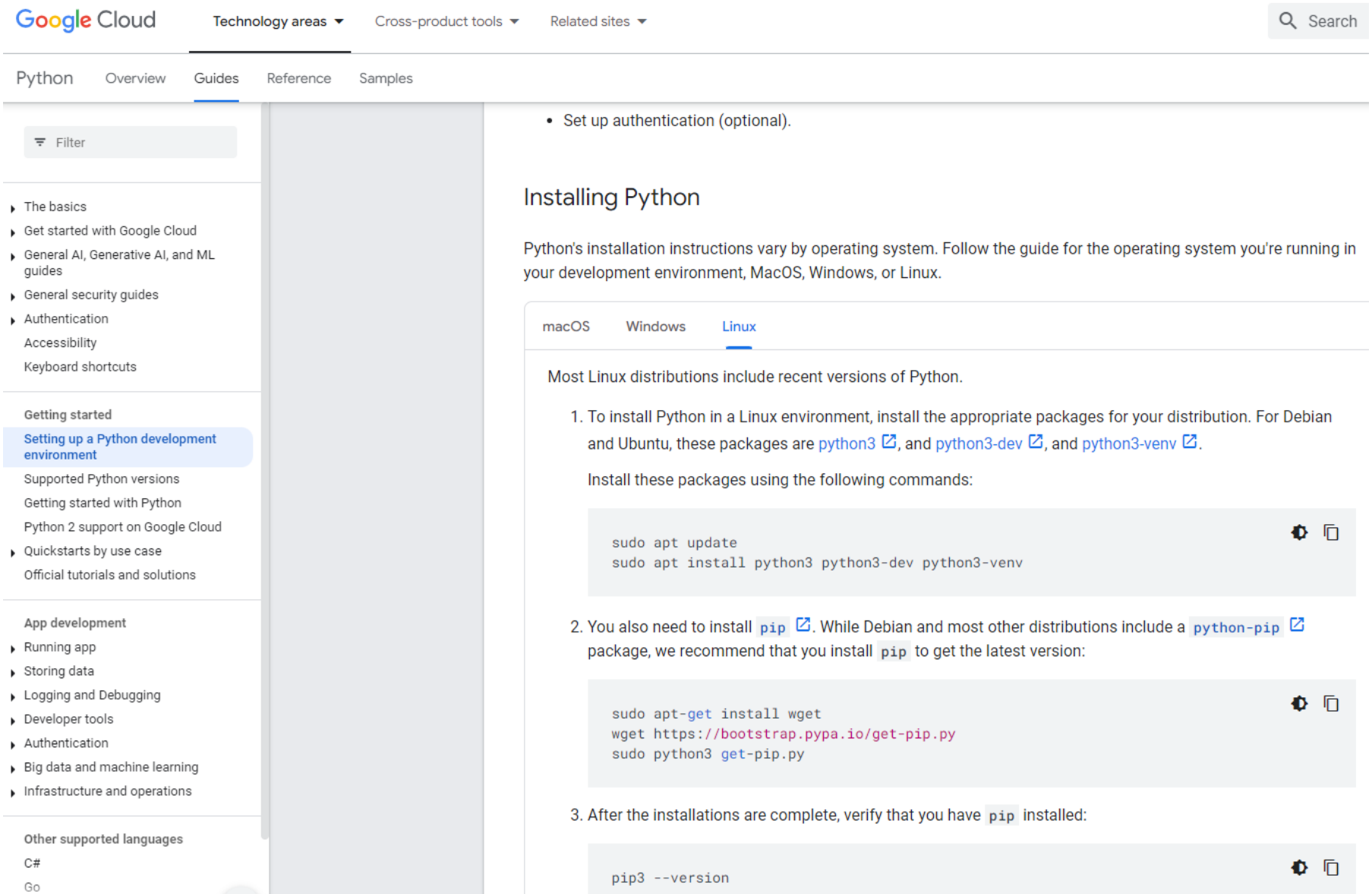
Google Cloud Console

gcloud – command line tool (possibly next lecture)

Compute Engine API



Instance Creation on GCP



Instance Creation on GCP: Installations

```
sudo apt update
sudo apt install python3 python3-dev python3-venv

sudo apt-get install wget
wget https://bootstrap.pypa.io/get-pip.py
sudo python3 get-pip.py

pip3 –version
```

Jupyter Notebook Setup

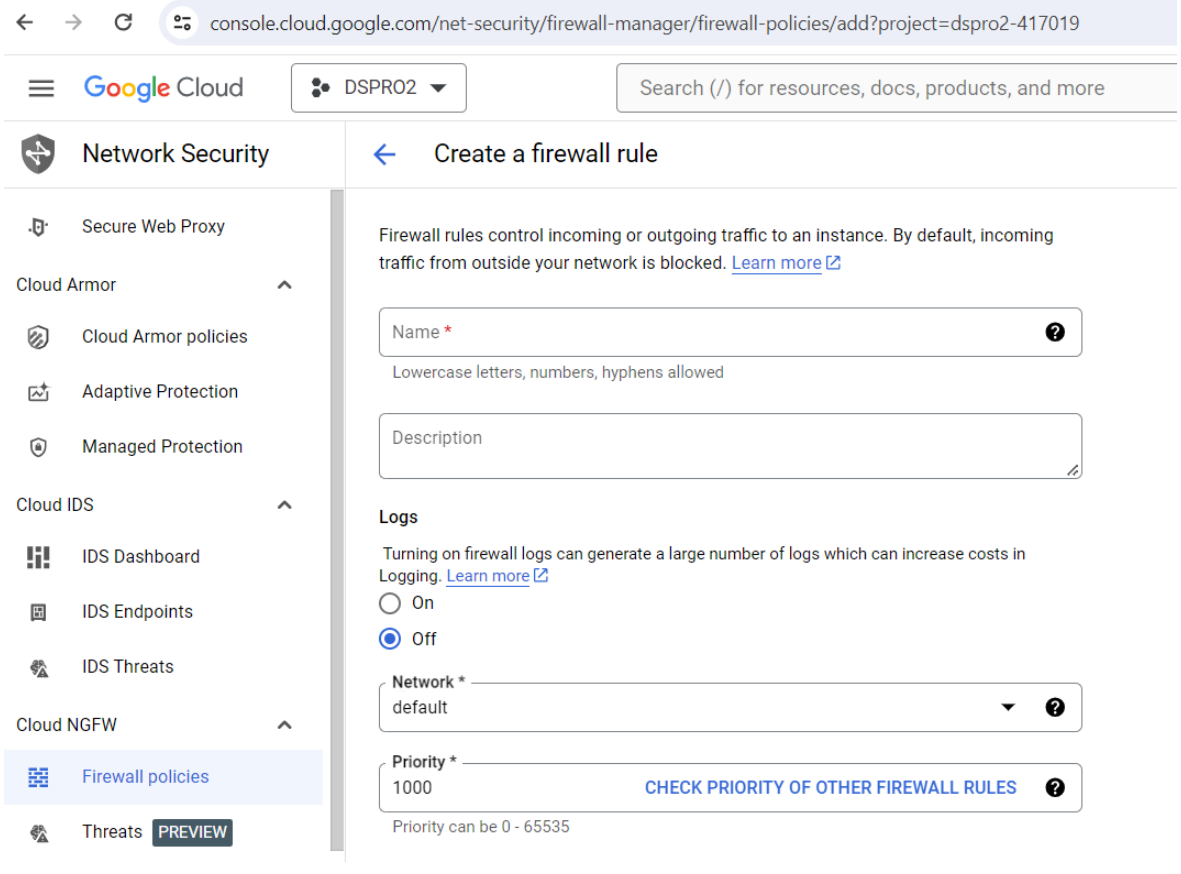
How to enable Jupyter Notebook Viewing in GCP VM?

3 steps:

Make external IP a static one

Create Firewall rules

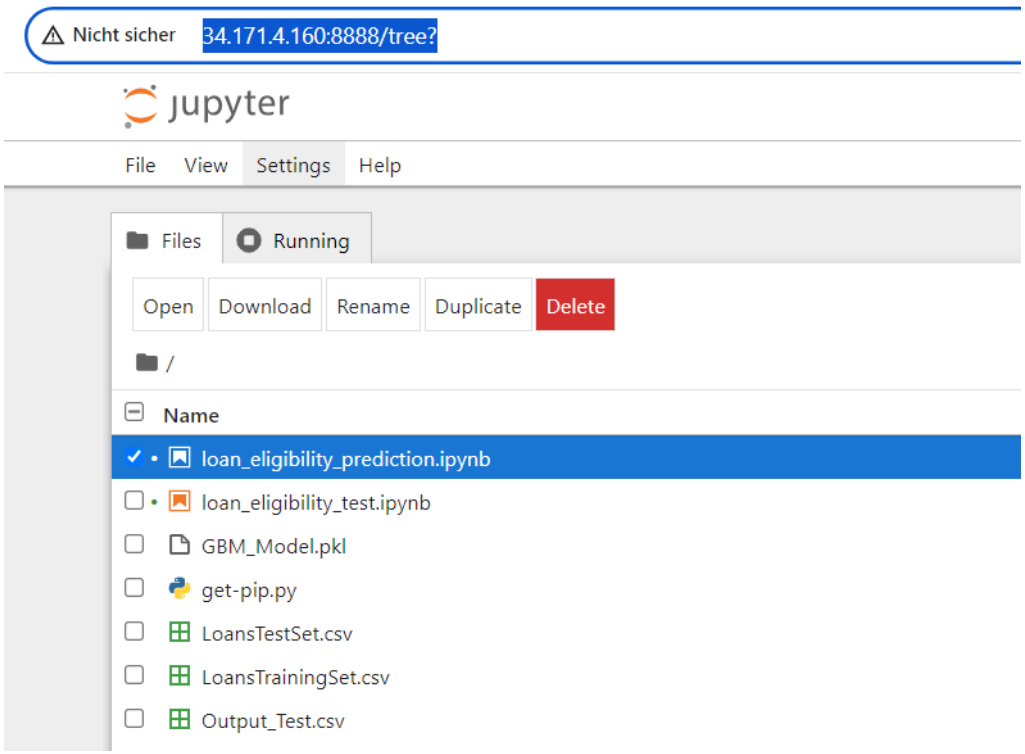
Jupyter notebook installations and configurations



Jupyter Notebook Installations

sudo apt install python3-pip
sudo pip3 install notebook
jupyter –version

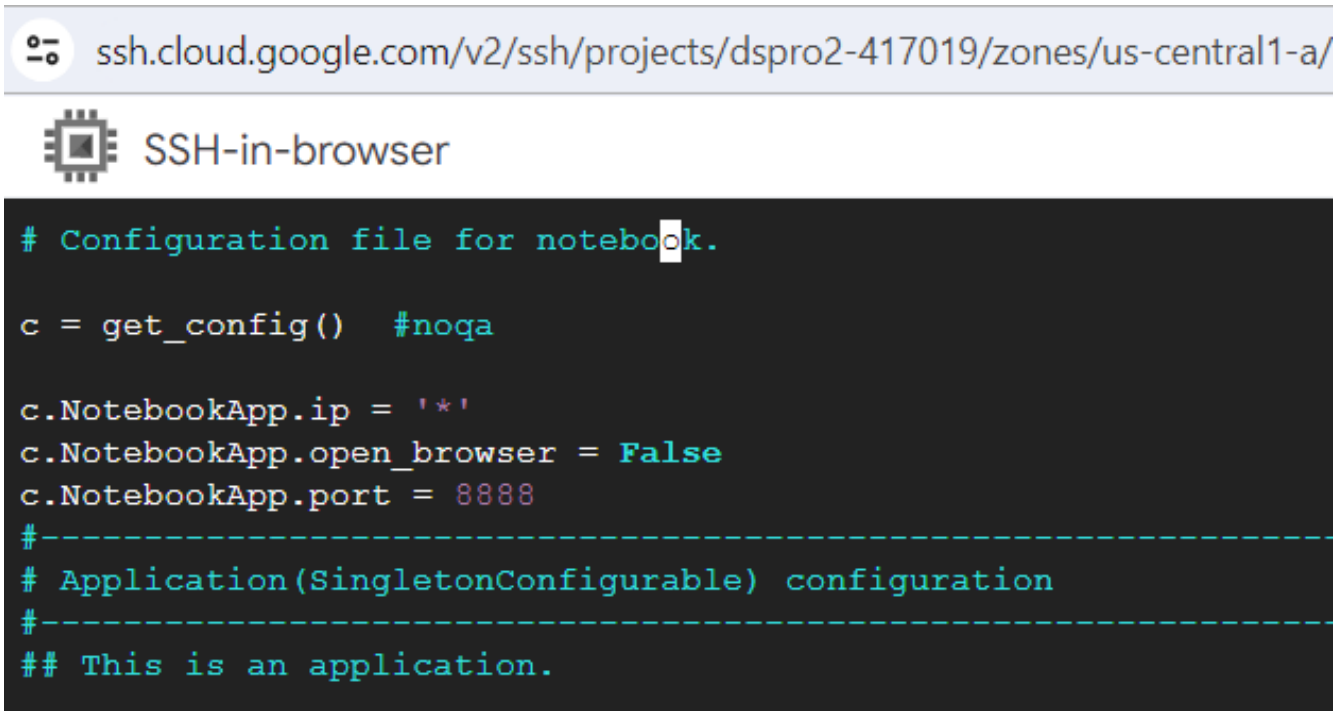
jupyter-notebook –no-browser –port=8888
http://External Static IP Address:Port Number
http://34.171.4.160:8888/



Jupyter Notebook Configurations

jupyter notebook –generate-config
vi /home/user name/.jupyter/jupyter notebook config.py

c.NotebookApp.ip = '*'
c.NotebookApp.open browser = False
c.NotebookApp.port = 8888



Library Installations

pip install –upgrade wheel numpy pandas matplotlib seaborn
pip install fancyimpute imblearn xgboost scikit-learn

```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import seaborn as sns
import statistics
from sklearn.model_selection import train_test_split,GridSearchCV,cross_val_score
from sklearn.preprocessing import LabelBinarizer,StandardScaler,OrdinalEncoder
from sklearn.metrics import confusion_matrix
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
from scipy.stats import boxcox
from sklearn.linear_model import LogisticRegression,RidgeClassifier, PassiveAggressiveClassifier
from sklearn import metrics
from sklearn import preprocessing
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from xgboost import plot_importance
from matplotlib import pyplot
from sklearn.naive_bayes import BernoulliNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import LinearSVC
from sklearn.tree import DecisionTreeClassifier
from xgboost import XGBClassifier
from fancyimpute import KNN, SoftImpute
from imblearn.over_sampling import SMOTE
import joblib
import operator
```