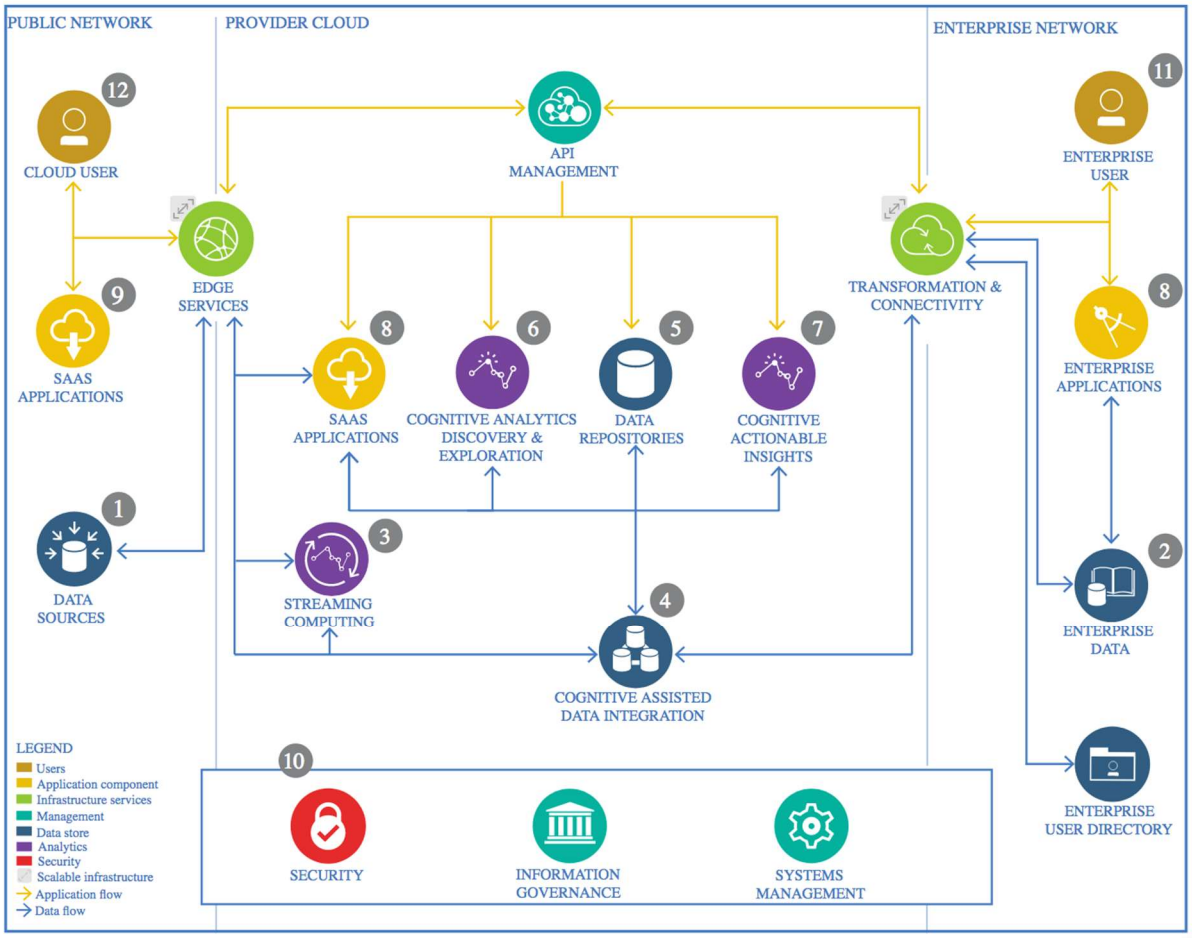
Forex Prediction using LSTM Neural Network

Ben Sung for Coursera Data Science Course

Architectural Decisions Document

In this project, I will do Forex prediction using LSTM using data of EURO vs USD from 2011 to 2018.

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

Data is the most important part of starting with the project for Data Science. The data used in this project was downloaded from QuantDataManager from Jan 2009 to July 2019.

This Forex Euro / USD dataset contains columns labels of Date, Time, Open, High, Low, Close and Volume.

1.1.2 Justification

QuantDataManager is one of the largest provider for Forex prices and popular among traders to get their online prices.

1.2 Enterprise Data

1.2.1 Technology Choice

There is no enterprise data used in this project.

1.3 Streaming analytics

No streaming analytics done in this project.

1.4 Data Integration

1.4.1 Technology Choice

Python with pandas, numpy, matplotlib, keras and h5py libraries
Popular platform for data scientist

LSTM from Keras using TensorFlow backend
Stateful deep learning algorithm suitable for time series prediction
Neural Network of 4 layers (Input, 2 LSTM layers with 10 neurons each and a Dense output layer)

Jupyter Lab
Easy collaboration, streamline and enable more productivity for data scientist projects

1.4.2 Justification

Python platform is the most used among data scientist.
For Time Series prediction, LSTM Stateful deep learning are normally used using Keras API to design the neural network layers with the parameters, epochs, batch size to do the training and testing/validation of the predicted data.

1.5 Data Repository

1.5.1 Technology Choice

Data for the Forex was downloaded from QuanDL website that is used by traders for multi currencies.

1.5.2 Justification

- I choose the previous Closed prices to predict the future prices, as they indicate the closed price of the day.
- Therefore for **feature engineering**, I drop the unrelated columns: Time, Open, High, Low and Volume.
- No other cleaning on the data necessary as there are no missing or error data.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Matplotlib library for Python is used in this project to plot or graph the data.
Pandas is used to manipulate Dataframe / Dataset.

1.6.2 Justification

Graphs/Plots are used for discovery and exploration of data.
Using Jupyter Notebook, we can explore the data to view and manipulate data by modifying the python code.

1.7 Actionable Insights

1.7.1 Technology Choice

LSTM neural network using Keras library

1.7.2 Justification

LSTM is known to be effective to predict time series data. It has three gates per LSTM cell (input, forget and output gates) and parameters of weights and bias to improve on the modeling accuracy.

1.8 Applications / Data Products

1.8.1 Technology Choice

Jupyter Lab

1.8.2 Justification

For predicting Forex prices a stand alone PC with Jupyter Notebook would be sufficient for now, as the data would not exceed millions of data. However, if more features or dimensions required to perform prediction with live streaming Forex data, then Apache Spark and clusters of machines are required.

1.9 Security, Information Governance and Systems Management

No security was required for this project.

1.9.1 Technology Choice

No security was required for this project.

1.9.2 Justification

No enterprise data are used in this project.