

Extra Virgin Olive Oil Price Prediction

Diego Hueltes

 @jdiegoh

<https://www.sli.do>

#bigdata2018

Jaén

- 20% of the Olive Oil global production
- 550.000 Ha. Olive trees
- Picual variety (high quality)
- 8M work days / campaign
- 300M€ / campaign







Production factors



Temperature



Rain



Extreme wind

Researches

CO2RBFN for short-term forecasting of the extra virgin olive oil price in the Spanish market

M.D. Pérez-Godoy, P. Pérez, A.J. Rivera, M.J. del Jesus, C.J. Carmona, M.P. Frías and M. Parras
Article in International journal of hybrid intelligent systems · January 2010

CO2RBFN for Short and Medium Term Forecasting of the Extra-Virgin Olive Oil Price

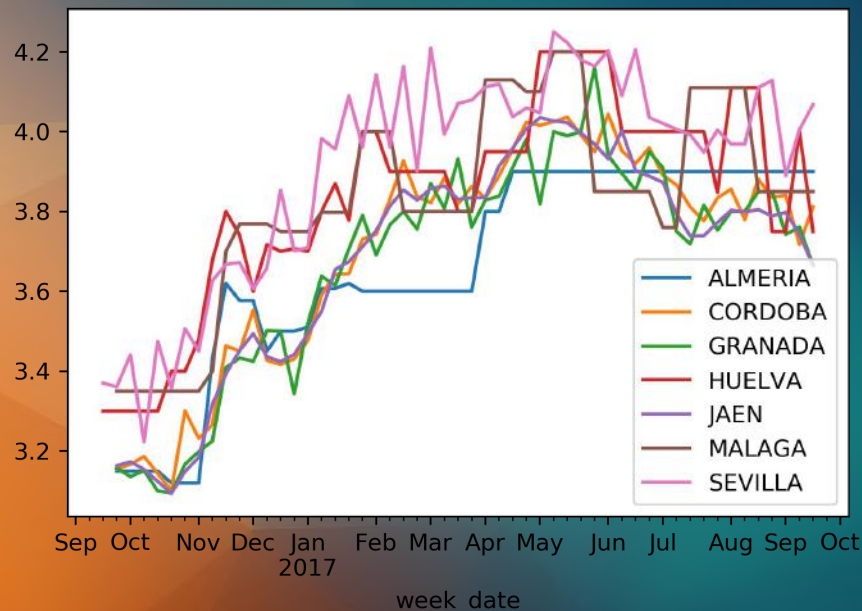
M.D. Pérez-Godoy, P. Pérez-Recuerda, María Pilar Frías, A.J. Rivera, C.J. Carmona, and Manuel Parras
Chapter · April 2010 DOI 10.1007/978-3-642-12538-6_10 ISSN 1860-949X

Data extraction - Prices

Extra Virgin Olive Oil price data
in origin, for Andalucía region

Weekly data, aggregated per
province

From Oct 2007 to July 2018

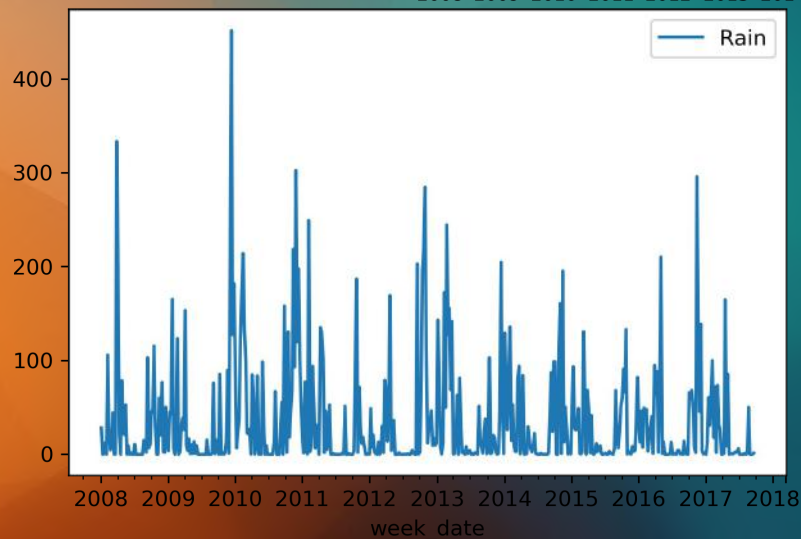
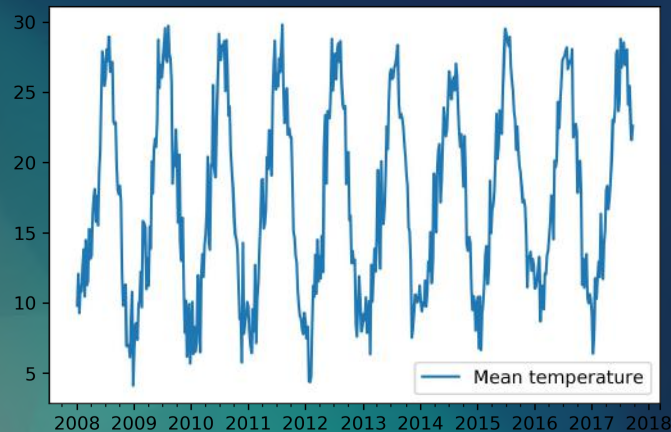


Data extraction - Weather data

Data provided by AEMET
(Spanish meteorology agency)

Daily data aggregated per city
but aggregated afterwards
weekly / province

From Oct 2006 to July 2018



Data extraction - Weather data

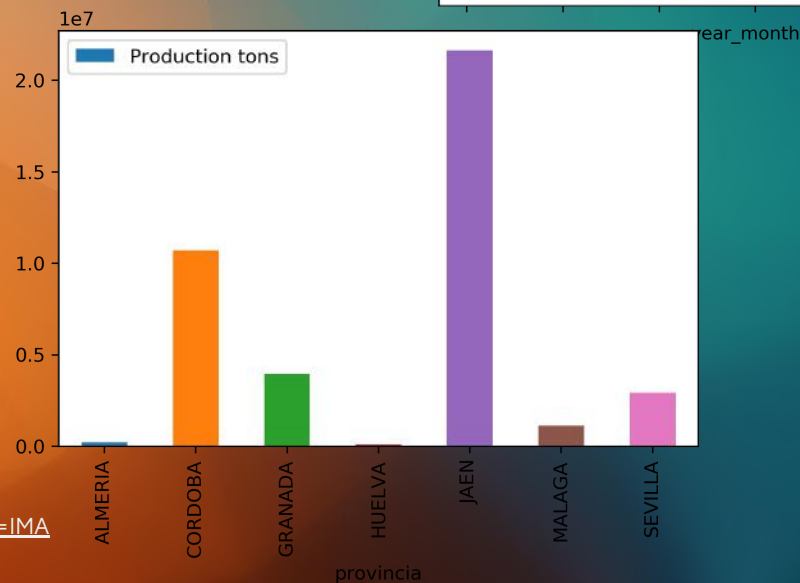
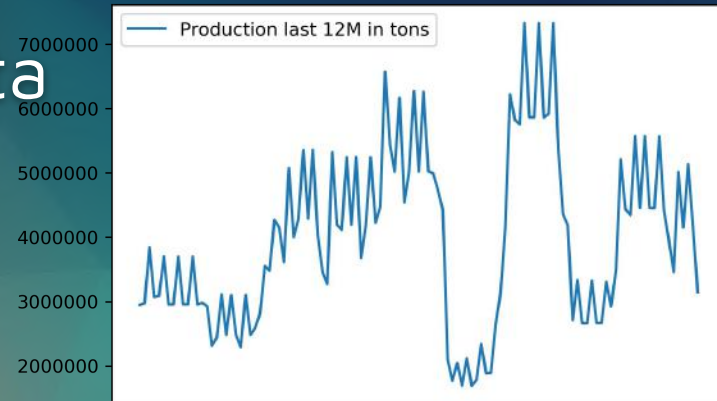


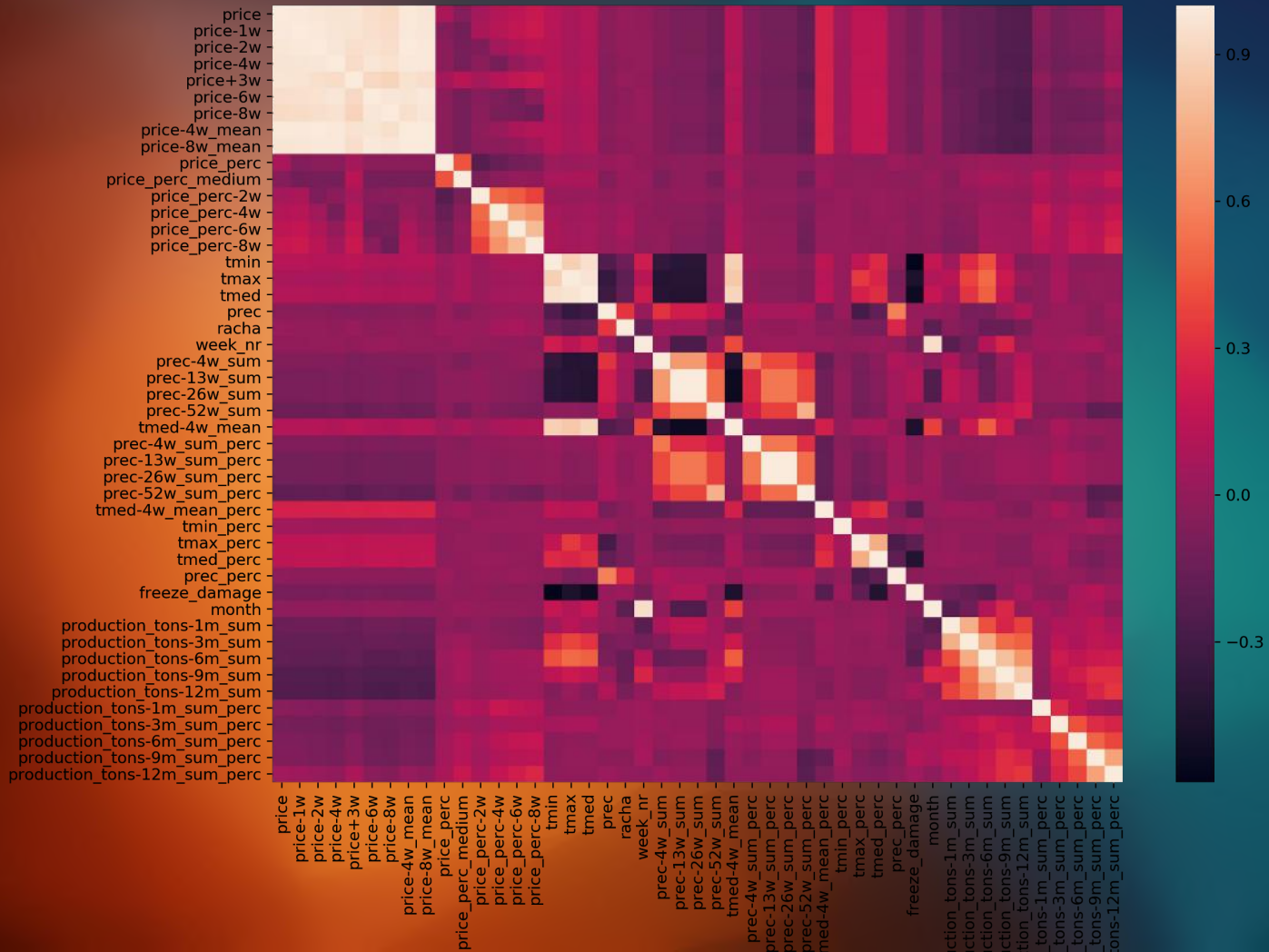
Data extraction - Production data

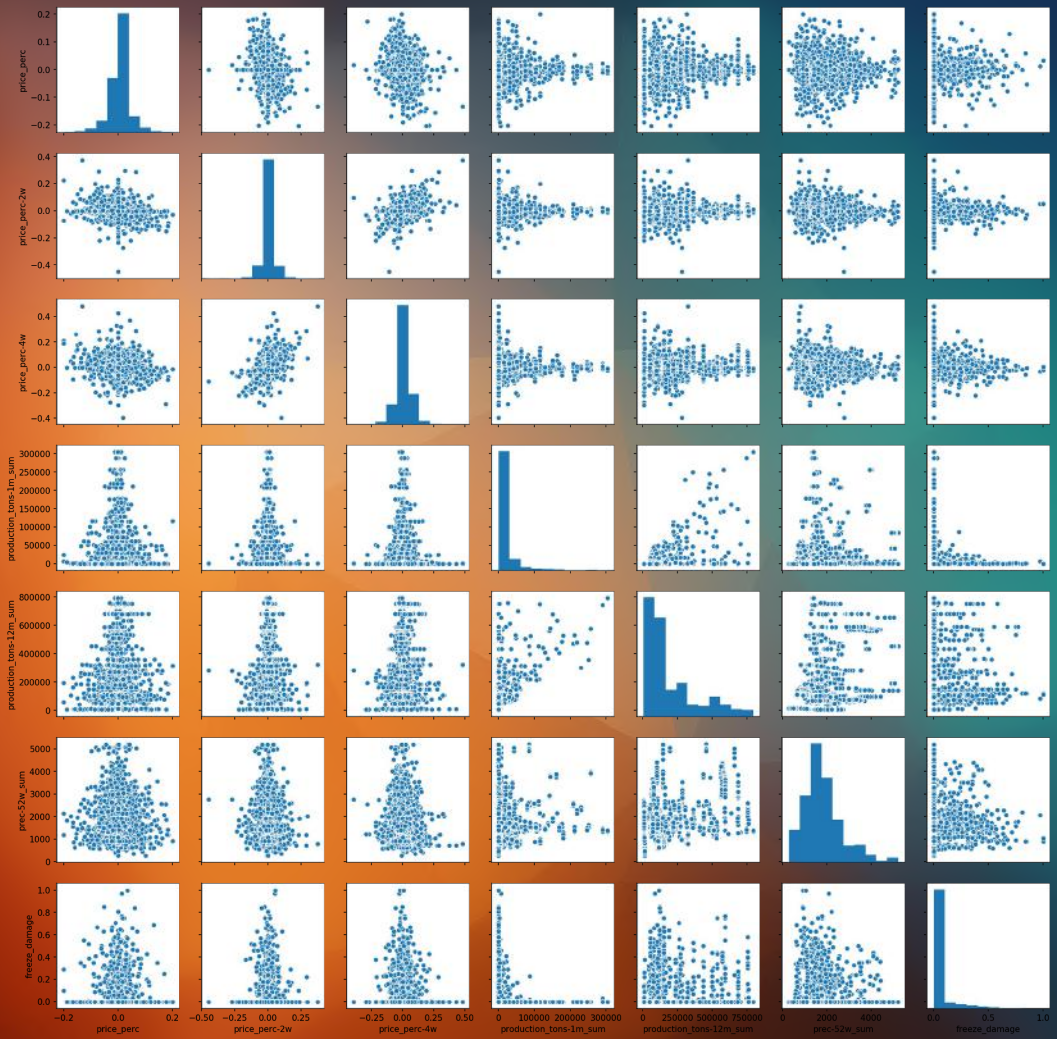
Data extracted from the Spanish department of agriculture.

Monthly data aggregated per province.

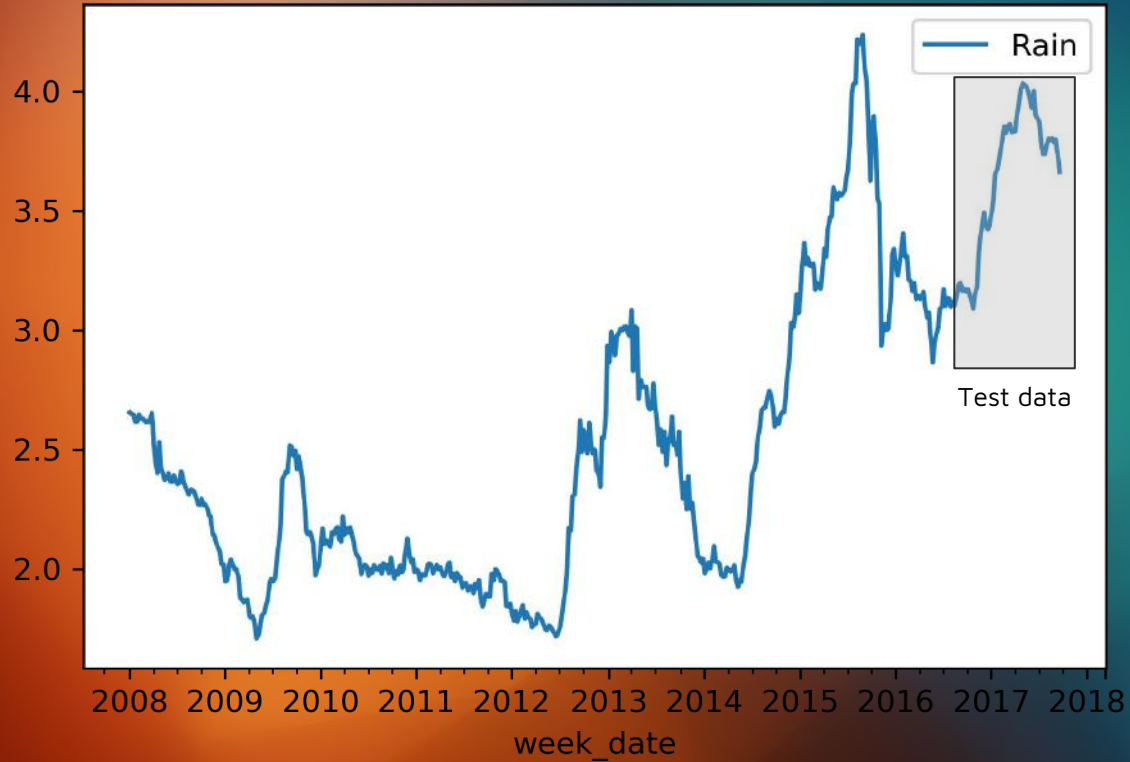
From Oct 2006 to July 2018



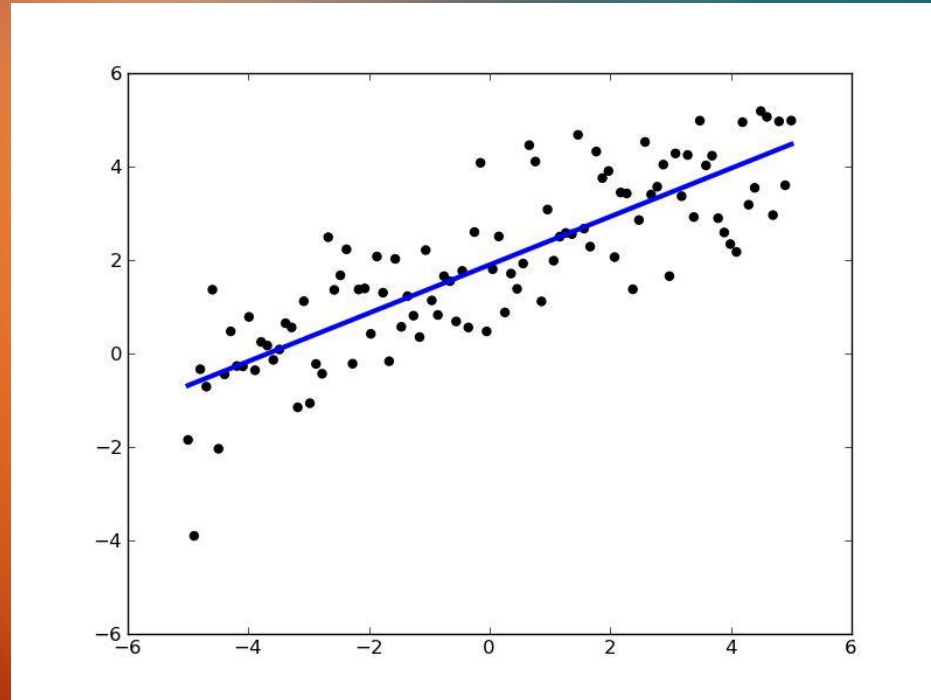




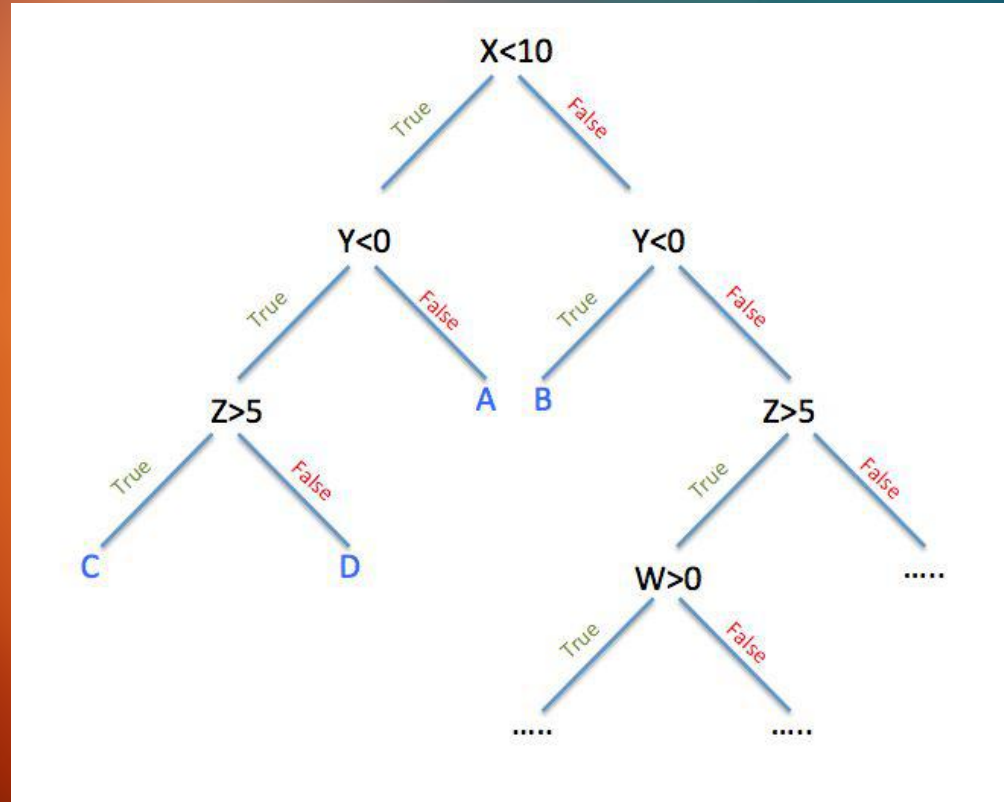
Train / test set



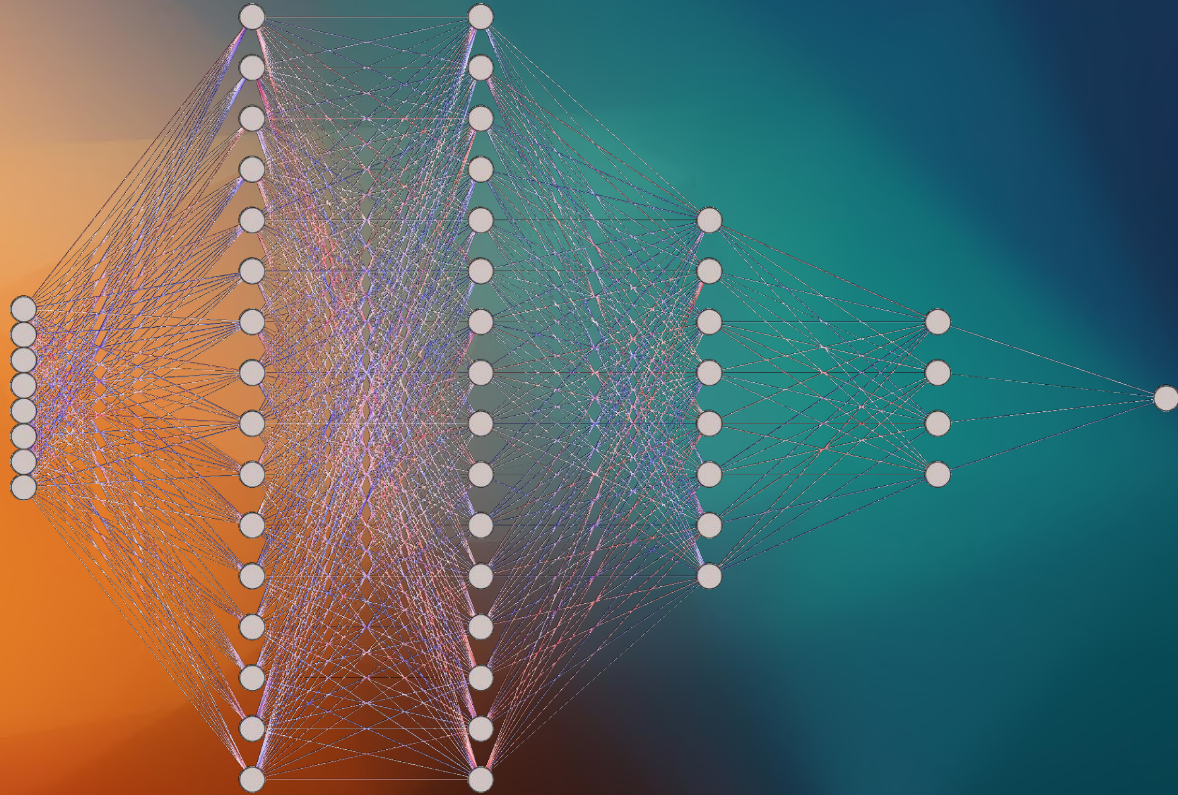
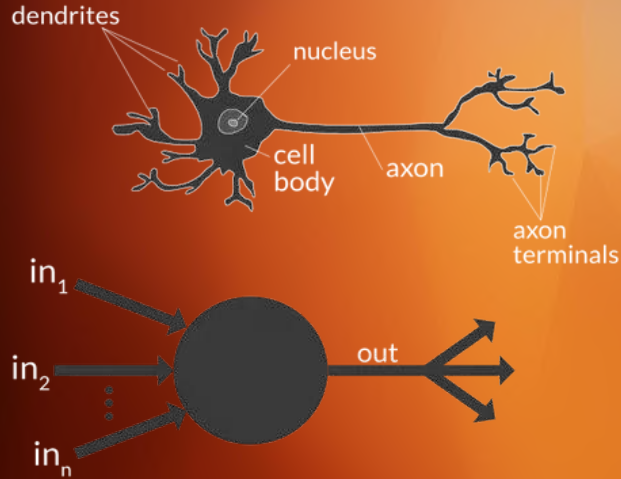
Prediction models: Linear models



Prediction models: Tree based models



Prediction models: Deep learning models



Short term prediction

Previous results from researches

NU - SVR: 1,936% MAPE

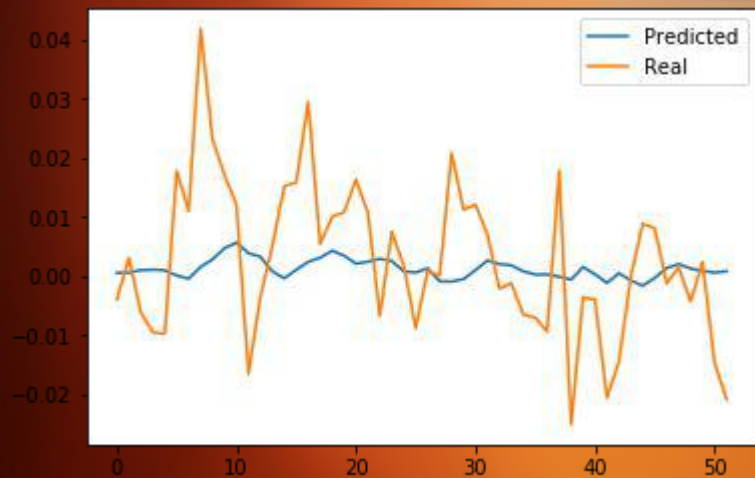
ARIMA: 1,303% MAPE

CO2RBFN: 1,914% MAPE

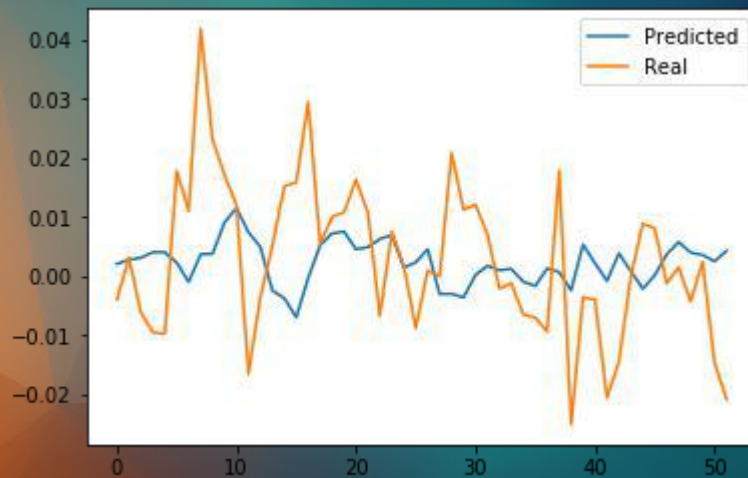
MAPE = Mean Absolute Percentage Error

Linear models

Lasso Lars

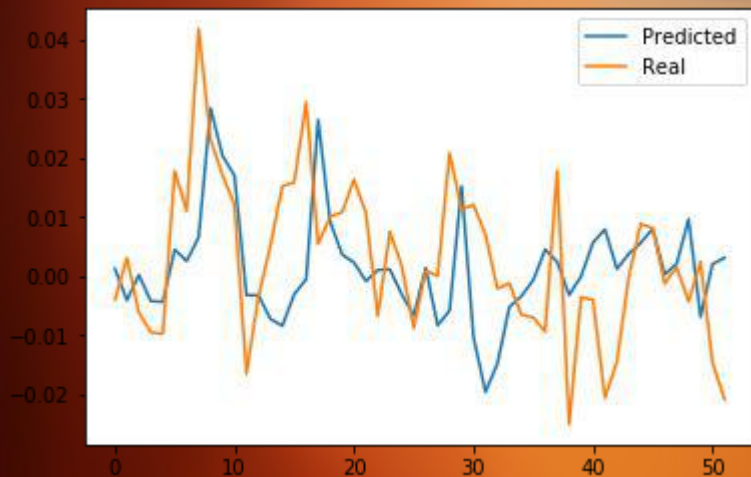


Linear Regression

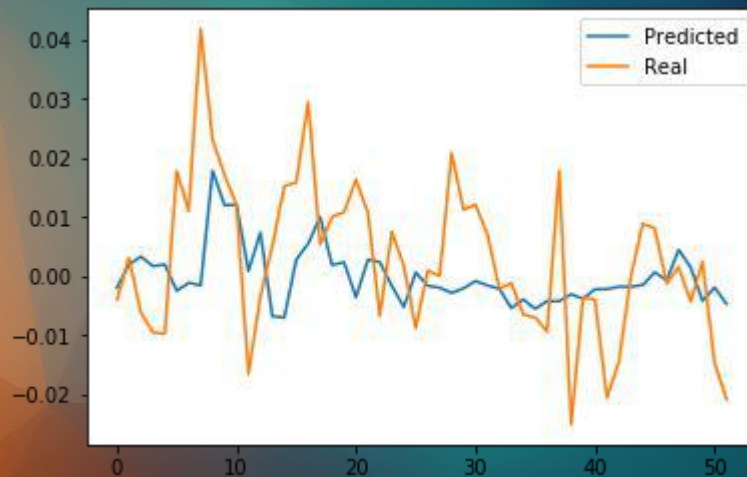


Tree based models

Random Forest



XGBRegression



Tree based models

Random Forest



1.090% MAPE

XGBRegression



0.998% MAPE

NVIDIA K80

- 4992 NVIDIA CUDA cores with a dual-GPU design
- Up to 2.91 teraflops double-precision performance with NVIDIA GPU Boost
- Up to 8.73 teraflops single-precision performance with NVIDIA GPU Boost
- 24 GB of GDDR5 memory





This notebook is thought to be executed in Google colab, but ignoring this first cell can be executed in any jupyter environment.

```
[ ] !pip install -U -q PyDrive
!apt-get install -y libhdf5-serial-dev
!pip install tables

from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials
import os

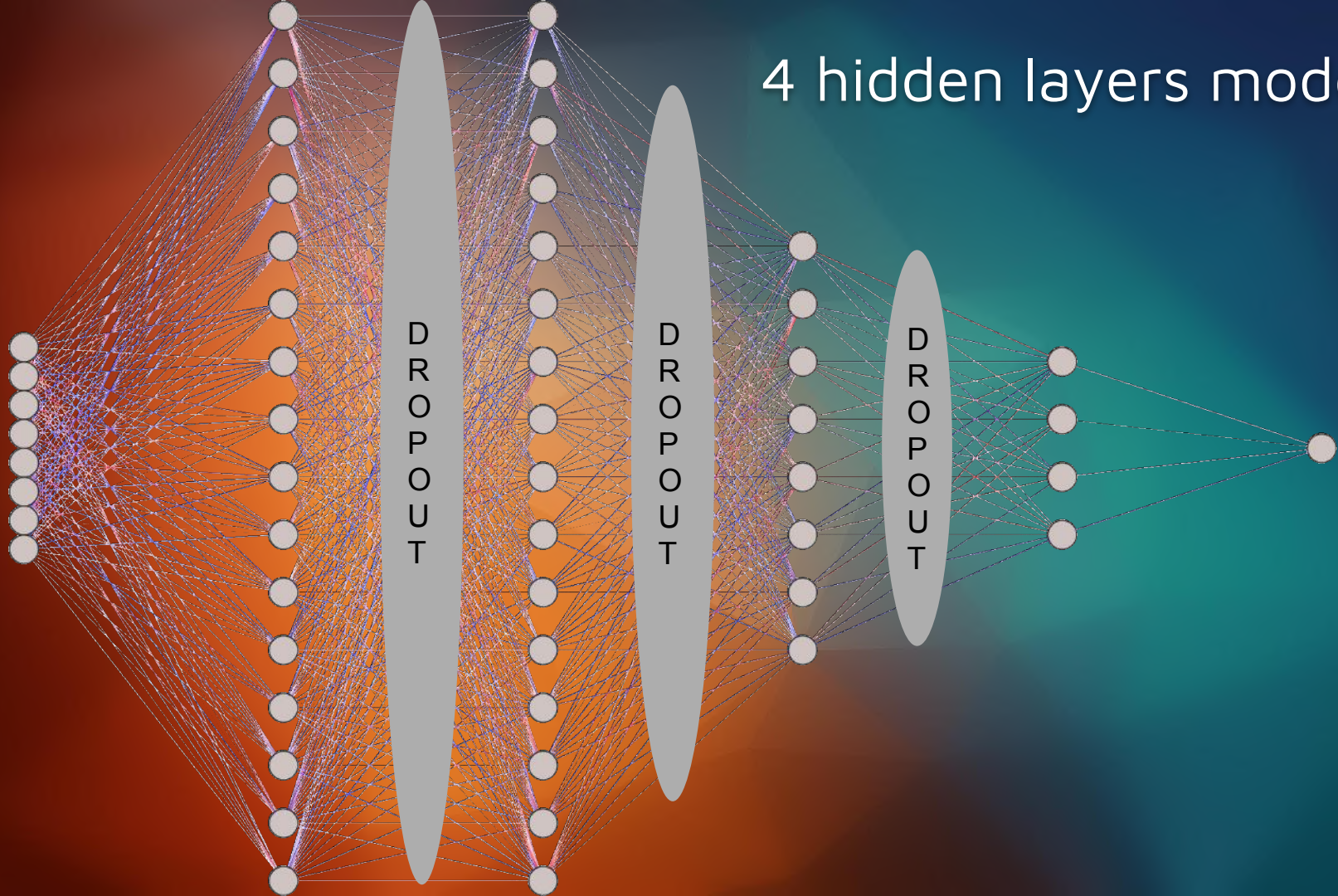
# 1. Authenticate and create the PyDrive client.
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)

local_download_path = os.path.expanduser('~/.data')
try:
    os.makedirs(local_download_path)
except: pass

file_list = drive.ListFile(
    {'q': "'11j5S_awb1RLtOH40jCcw7zIzRUSo0SS3' in parents'}).GetList()

for f in file_list:
    print('title: %s, id: %s' % (f['title'], f['id']))
    fname = os.path.join(local_download_path, f['title'])
    print('downloading to {}'.format(fname))
    f_ = drive.CreateFile({'id': f['id']})
    f_.GetContentFile(fname)
```

4 hidden layers model

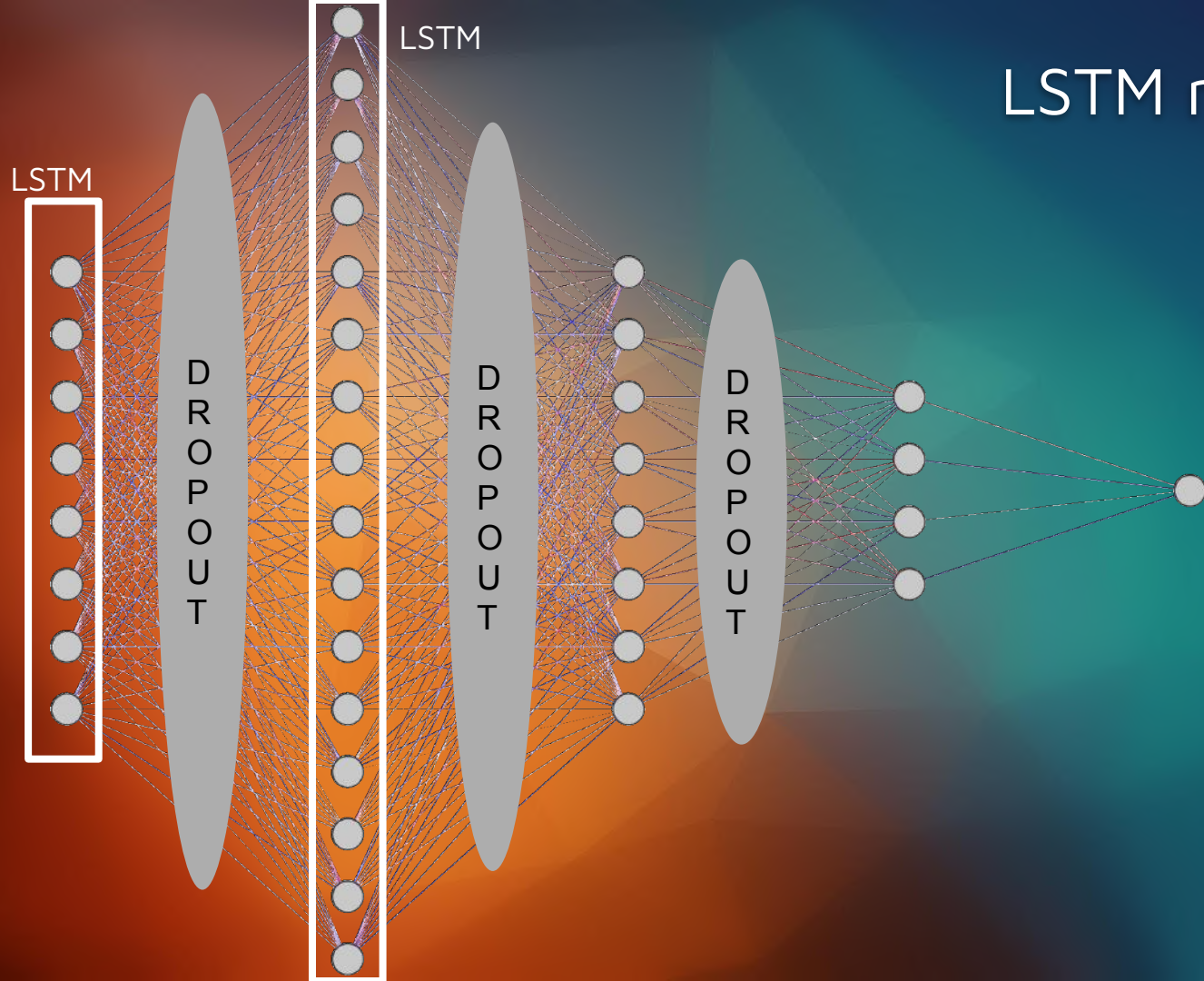


Deep learning model - 4 hidden layers

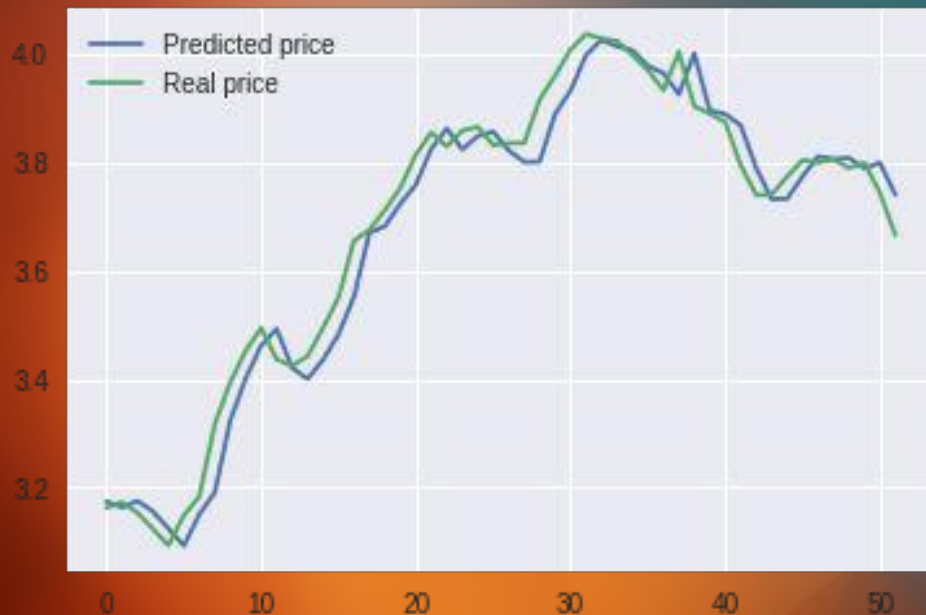


0.982% MAPE

LSTM model



Deep learning model - LSTM



1.073% MAPE

Medium term prediction

Previous results from researches

MLP ConjGrad: 2,970% MAPE

NU-SVR: 3,003% MAPE

CO2RBFN: 3,230% MAPE

MAPE = Mean Absolute Percentage Error

Deep learning model - 4 hidden layers



3.320% MAPE

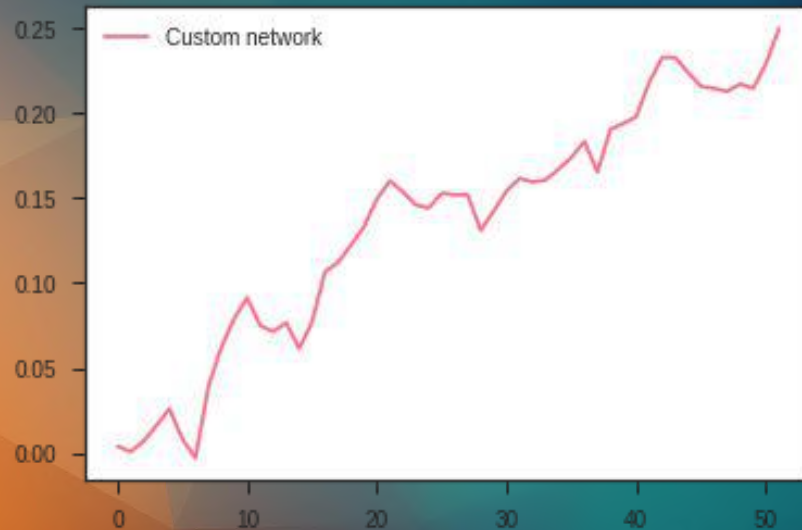
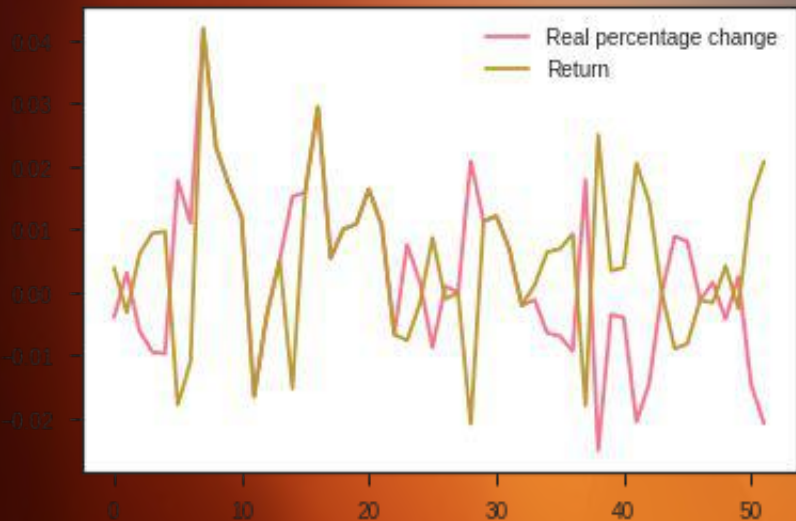
Linear (LassoLars) model



2.842% MAPE

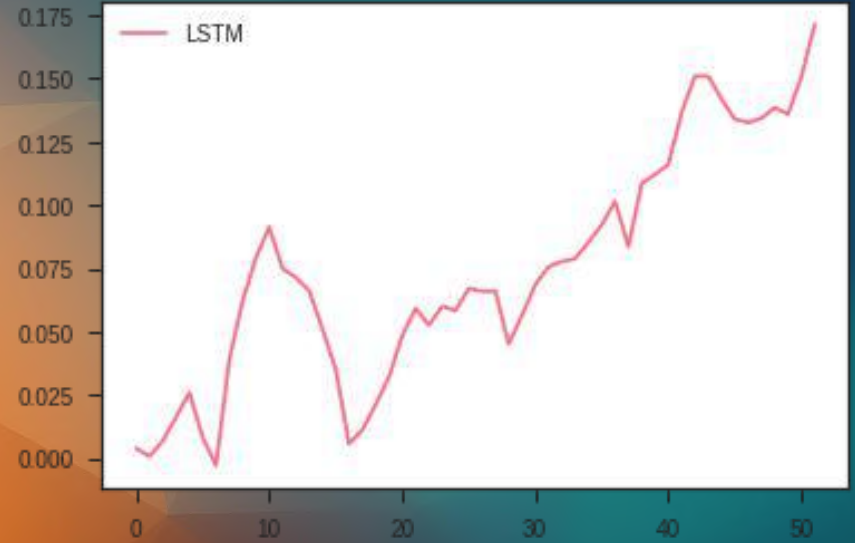
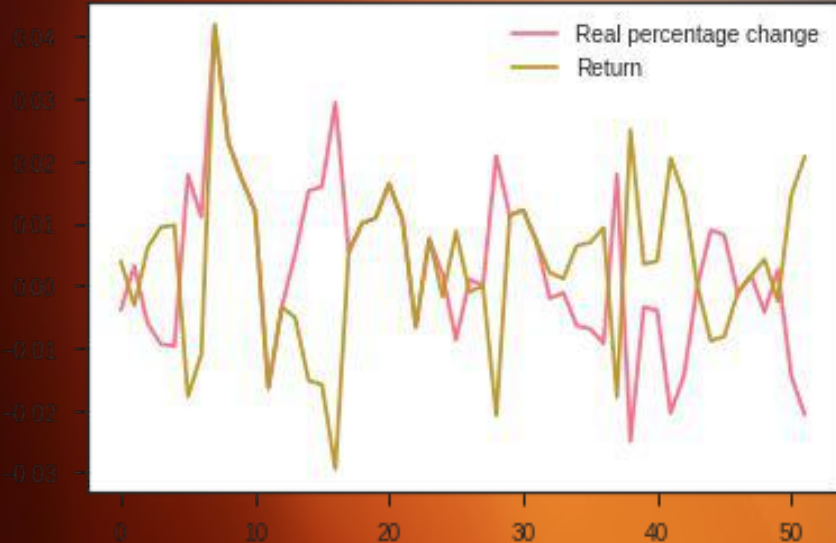
Short term price
direction prediction

Deep learning 1 hidden layer



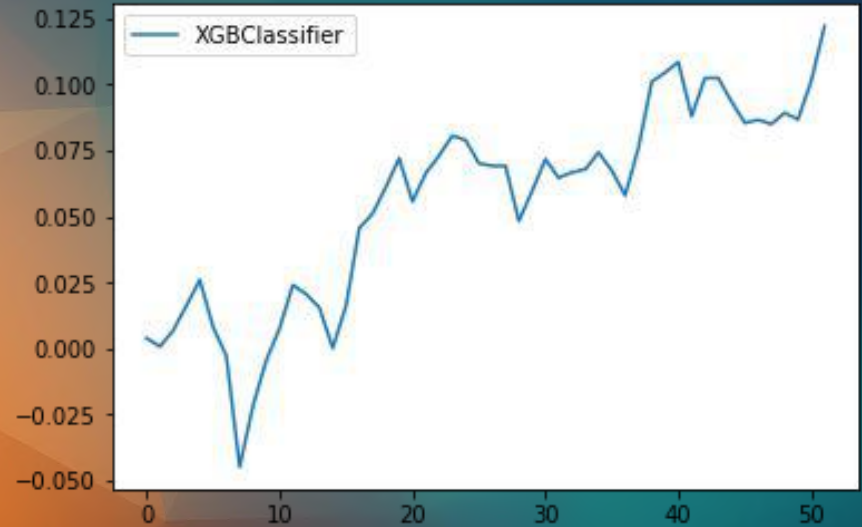
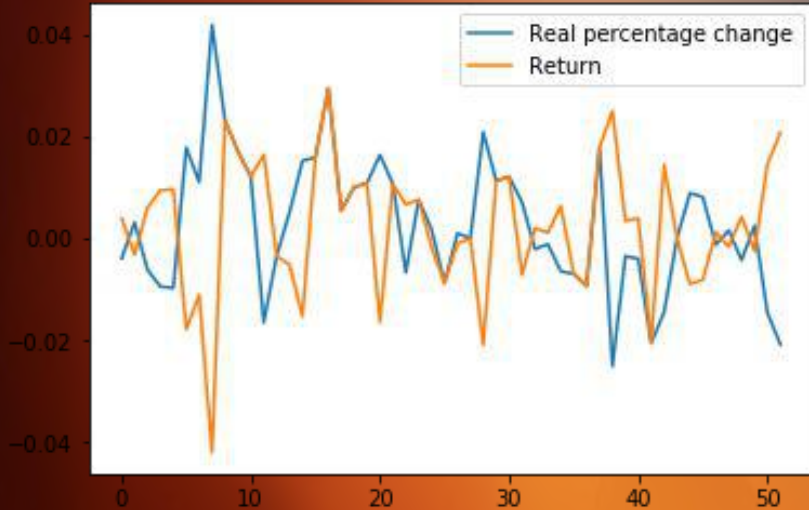
61% Accuracy 25% annual return

Deep learning LSTM 2 hidden layers



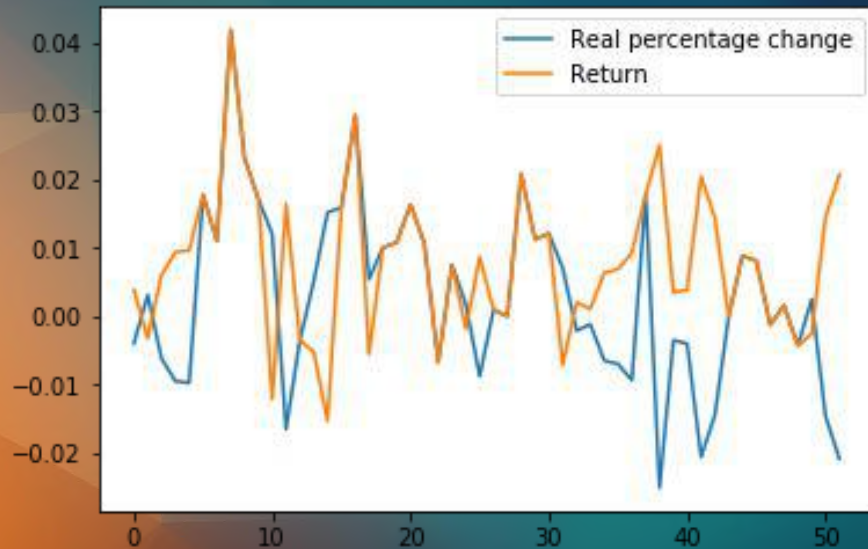
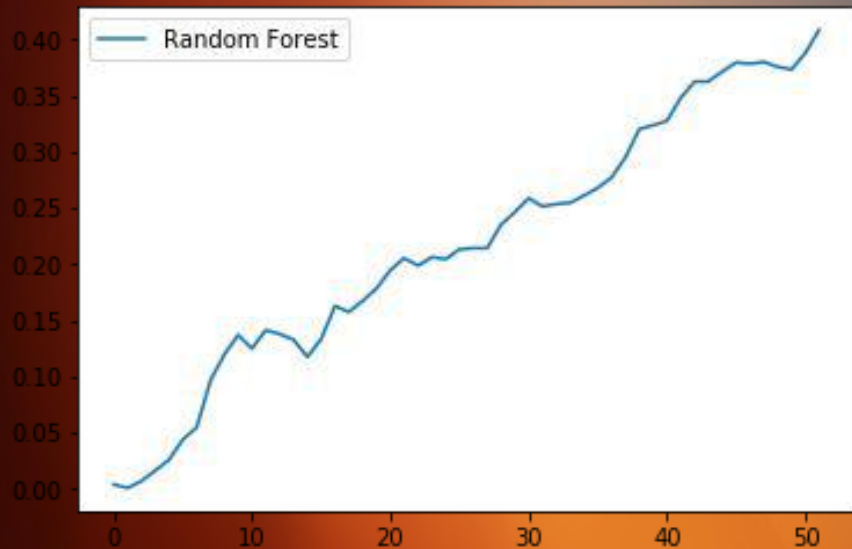
62% Accuracy 17% annual return

XGBoost Classifier



62% Accuracy 12% annual return

Random Forest Classifier



76% Accuracy 40% annual return

Improvements

- Daily data
- Full country data
- International market data
- Imports / Exports
- Other varieties of Olive Oil

Thank you!



@jdiegoh



DiegoHueltes



Diego Hueltes



diego@hueltes.com



<https://www.hueltes.com/olive-oil>