I have identified some 4-6 oil crisis starting dates [OIL-CRISIS-DATES] during 1990-2019, Figure 1.

It is a highly skewed time-lapse data, considering that the total number of observations in the oil price series (monthly, 1990-2015 equals 12x(2015-1990)=12x25=300 data points x).

Below is my outline of how to deal with a highly skewed data.

Highly skewed datasets and how to achieve high prediction performance on the minority class.

1. **imbalanced classification: Cost-sensitive supervised learning algorithms**
   Given labeled examples from the minority (rare) class, and it tries to improve prediction performance especially on the minority class.
   - 1.1. **Cost-sensitive KNN (C-KNN),** assigning class-based weights to each instance, thus weighing the votes of the neighbors.
   - **1.2. Cost-sensitive SVM** most commonly used methods in imbalanced classification.
   - 1.3. **Cost-sensitive Logistic Regression**
   - 1.4. **Performance comparison s.t. F-1 scores**
     compare the performances of cost-sensitive k-NN, cost-sensitive SVM, and cost-sensitive logistic regression
   - **1.5. Multiple Kernel Learning for Imbalanced Classification**

2. Unsupervised Anomaly Detection a.k.a. **rare class detection**

   Aims to detect the rare classes de-novo from a few examples.

   Detection of rare categories when no labeled samples are available a-priori

   Unsupervised Anomaly Detection - often unsupervised (or labels are known only for the “normal” data).

   2.1. three separate classification experiments:

      2.1.1. **Positive Compact**

      assumes that the minority class is compact and learns only from the minority class (with nNegative instances from the majority class added in the training set),

      2.1.2. **Negative Compact**

      the second experiment assumes that the majority class is compact and learns only from the majority class (with nPositive instances from the minority class added in the training set).

      2.1.3. **one-class SVM.**

      use all data, regardless of their labels.

   2.2. Results analysis as measure the F-1 score of Positive Compact, Negative Compact, one-class SVM and cost-sensitive logistic regression on highly skewed datasets

      2.2.1. All experiments use Gaussian RBF kernel / Python Scikit-learn one-class SVM classifier

      2.2.2. Multiple Kernel Learning approach with compactness assumption

         2.2.2.1. Compare the MKL approach to random under sam-pling (RUS), one-class SVM, and cost-sensitive logistic regression

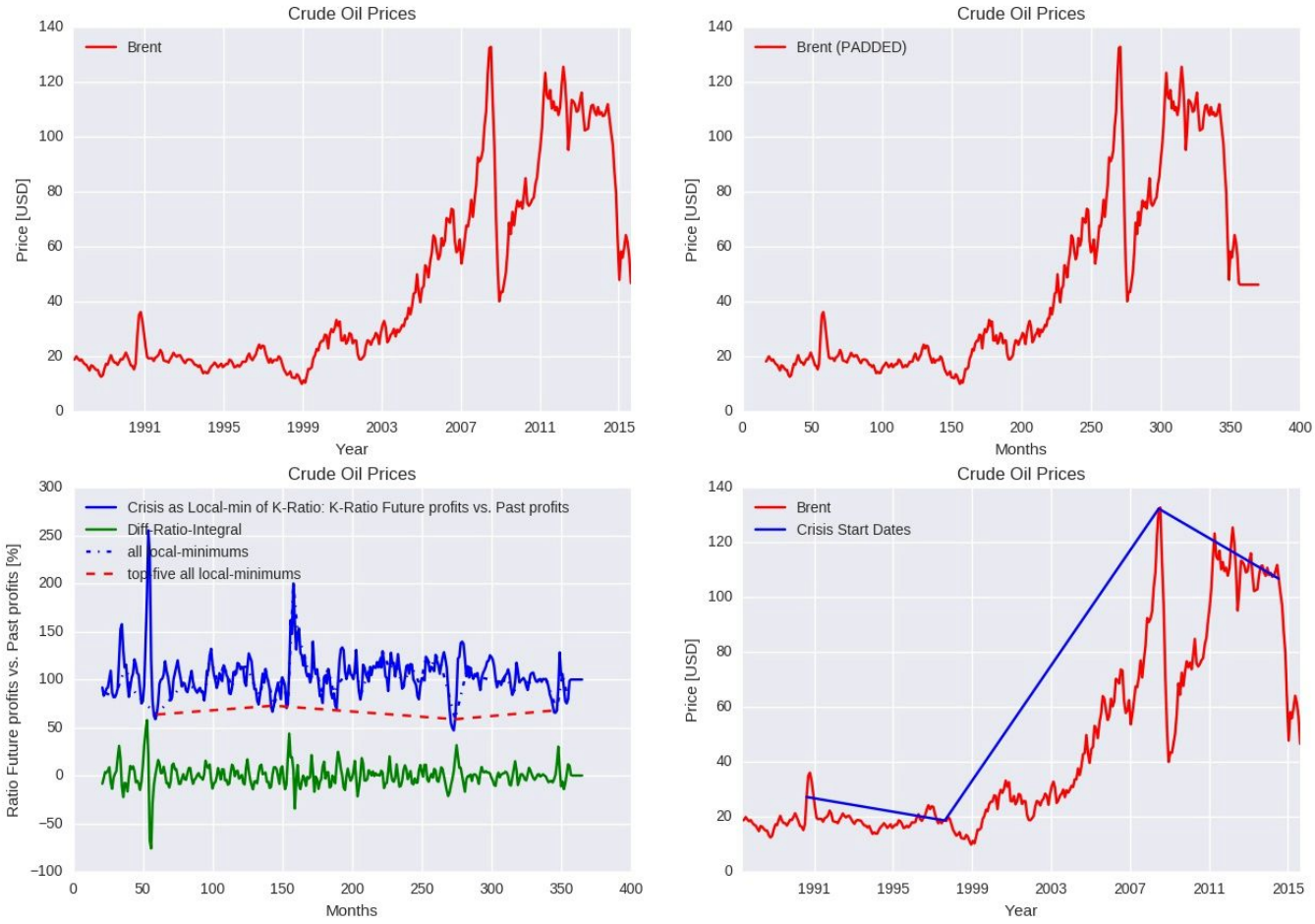

Figure 1
Top Left: Brent oil price
Top Right: Brent oil price extended by 2 years (padding)
Bottom Left: K-ratio function (blue), its minimum (dashed) zero-crosses of 1st derivative (green)
Bottom Right: The four Crisis-Dates (blue) over Brent oil price

K-ratio is computed at each time as ratio:

accumulated oil price per barrel for 12 months ahead vs.
accumulated oil price per barrel for 12 months in the past