
COMMODITY PRICE FORECASTING

MLND CONNECT FEB '18

VIBHAV LAL
JUNE 2, 2018

DOMAIN BACKGROUND

- Commodity is one of the largest traded (by volume) commodity Internationally as well as domestic.
- Commodity markets are highly fragmented pricing.
- No standard price, which is why there is no exchange for this product.
- No set price for any product, production cost differs from producer to producer.
- Multitude of categories of products/items/sizes/grades within the categories with substantial price differences.

PROBLEM STATEMENT

- Use Machine Learning concepts to predict pricing fluctuations for a particular commodity item
- Determine significant correlations between various features effecting the commodity pricing.
- Implement an interactive commodity price forecasting mechanism classifying them under the following categories:
 - Specific product
 - Specific region
 - Time range

CHALLENGES WITH COMMODITIES

- Very fragmented buying-selling process – in the form of brokers, agents, traders, consumers, stock-holders, producers.
- Unorganized and unregulated – no need for any special license or membership is required.
- Buying methods are still very old-fashioned.
- Highly relationship based.
- Price intelligence – mostly through verbal communication.
- Published journals give historical prices and informal projections but no specific price predictions.

FEATURES CONSIDERED

- **Factors influencing Commodity prices:**

- Production/availability/demand
- Imports
- Exports
- Iron pricing
- Coal pricing
- Petroleum prices
- Construction
- Infrastructure projects
- Bank financing
- Interest rates
- Currency fluctuations
- Government policies
- Weather conditions
- Labor market and policies
- Ocean freight
- Logistics
- Seasonality

DATASETS AND INPUTS

- **Data Sources:**
 - Industry journals
 - IMF Primary Commodity Prices
 - Department of Commerce – Data catalog (data.gov)
 - Trade sources (informal)
 - Import/export data from Customs
 - Historical Commodity Prices (datahub.io/dataset/commodity-prices)

PROPOSED SOLUTION

- Determine correlation between various variables in order to predict a pricing trend of the commodity.
- Use the correlation matrix and Time Series Forecasting to achieve a high accuracy on the commodity pricing in the following categories:
 - Time range
 - Region specific
 - Product category

BENCHMARK MODEL & EVALUATION METRICS

- Correlations generated results will be compared against historical data
- Achieve high accuracy of the predicted prices compared of historical data
- Attempt to predict pricing slightly better than London Metal Exchange

PROJECT DESIGN

- Importing and preprocessing the data
 - Merge several dataset for a particular commodity from different sources
 - Commodity monthly dataset is available from 1980-2017
- Preparing training and test data
 - Training & testing data split chronologically 80:20
- Data scaling
 - Scaling in Python using sklearn MinMaxScaler
 - Activation functions to scale the data available
- Techniques
 - Apply Naives Bayes & Bayesian Models
 - Apply Classifiers & Logical regression to find correlations between several impacting factors
 - Convolution neural networks to apply differing weights for each of the features