Chapter 6

Conclusions & Recommendations

6.1 Introduction

Over the past few decades, a rich set of data analysis techniques, including methods and algorithms, are been applied to select relevant features for oil price forecasting. There have been enormous development in data mining community in recent years but most of the studies related to oil price predictions have been concentrated in developing econometric time series models. This thesis has dealt with the challenging problem and tries to propose $MI^2$ and $I^2MI^2$ algorithm to provide the set of relevant and non-redundant features that can increase oil price forecasting performance. The effectiveness of the proposed algorithm with neural network as forecasting engine has been examined and evaluated with several state-of-art feature selection methods.

Section 6.2 discusses conclusions based on data analysis and evaluations. Section 6.3 provides certain recommendations for future researchers. The limitations of this study are discussed in section 6.4. Section 6.5 provides an overview of general contribution of this study in field of data mining and applied energy. Section 6.6 discusses the future scope of this study for researchers.
6.2 Conclusions

Based on the research study and analysis the following conclusions have been reached.

- **Impact of emerging economies**
  The proposed algorithm proved the importance of emerging economies (Non-OECD Consumption, China Consumption and China Reserves) in driving oil prices. The results from both groups identified China Reserves as a key variable for deriving the future price path. The proposed algorithm has performed well in figuring out the new changes in relationship between WTI and various factors. China consumption and its reserves have emerged as influential factors driving oil prices post 2008 financial crisis with drastic increase in their respective percentage contributions.

- **Recent change in impact of Non-OECD consumption compared to OECD Consumption as influential factor**
  According to British petroleum [5], Non-OECD consumption grew by 5.3% in track with 10-year average. It is evident from recent reports from EIA or BP that OECD Consumption tends to fall, while Non-OECD Consumption is projected to increase. The proposed algorithm selects Non-OECD Consumption as an key factor in driving oil prices. This confirms the superiority of the proposed algorithm in figuring out the recent change in data and identifying most relevant and non-redundant features for the study.

- **Non-OPEC Production & OPEC Supply as an emerging factor driving oil prices**
  Periodically, fluctuations in oil prices have been repercussion of OPEC news regarding cuts in production targets or changes in OPEC policies. OPEC announcements regarding change in policies or shift in production targets lead to change in oil prices. The impact of OPEC Supply still dominates the fluctuations in oil prices and the proposed methodology works well in identifying such relevant indicators driving oil prices. With recent change incorporated by importers to become exporters, U.S and China strive to optimize their domestic resources and become self-sustained to meet their own requirements. With increase in Non-OPEC production, the influence of older giants (OPEC) is diminishing as the most influential factors driving
oil prices. The proposed algorithm is effective in finding the most relevant features for the forecasting of oil prices.

- **NYMEX future prices is not a sole indicator**
  Many institutions—including central banks and international organizations—are currently using NYMEX future prices as a key indicator for deriving the directions of spot prices. The results have shown there are number of external factors, which are driving crude oil prices. The explanatory power of NYMEX future prices is around 16% whereas 84% accounts for other factors that influenced oil prices over 17-year time period (result obtained through $MI^3$ algorithm + GRNNN based methodology). Further, the explanatory power of NYMEX future prices is around 62% while 38% share is explained by external factors (result obtained through $I^2MI^2$ + GRNNN based methodology). The results proved the importance of identifying key factors for deriving the future path of oil prices and not to focus on a single indicator.

- **Petrodollar Effect**
  U.S Dollar Index has become the main factor driving oil prices, its percentage contribution to WTI price fluctuations is around 17% of total share. U.S dollar is the basic reserve currency and more than 80% of all international currency transactions involves dollar. Oil is traded in global market and most of the trade has operated and continue to operate in dollars, even if U.S is not the trade partner. Since oil prices are defined in terms of dollars by most oil exporters, and as a result, oil importing countries also pay in dollars. This petrodollar cycle is an important factors which is highlighted by $I^2MI^2$ algorithm based methodology.

- **Effect of CPI and EPPI in influencing oil prices**
  Periodically, the relationship between CPI, as measure of inflation, and oil prices has been causal. Oil currently provides the majority of human energy requirements. Economy of any country can’t run without oil and any fluctuations in oil prices effects economy directly. When oil prices are high, it leads to slower economic growth. High oil prices leads to high inflation initiating slower economic growth. The percentage contribution of CPI in influencing oil prices is around 37%. EPPI reflects the change of energy market and thus, the percentage
contribution of EPPI has been 25% for 17-year time period. CPI has the largest contribution followed by EPPI and DER as key drivers of oil prices.

- **Role of Speculation and Reserves before Crisis**
  Before the crisis, the effect of speculation in deriving oil prices increases with upward trend and shows strong explanatory power. After the crisis, its position weakens owing to high risk in crude oil markets and traders becoming susceptible of investing in oil markets. The proposed methodology is able to identify shifts in factors driving oil prices before and after 2008 financial crisis with high explanatory power. The role of reserves before the crisis seems to be enormous but weakens after the crisis. The importance of reserves before the crisis was repercussion of cuts in OPEC production targets or changes in OPEC policies. But after the crisis, increase in Non-OPEC Production indirectly weakens the effect of reserves on oil prices.

- **Original mechanism broke due to 2008 financial crisis**
  The overall mechanism of oil market broke after the crisis with EPPI, DJI, CC and CR being the influential factors driving oil prices. These four variables define the minimal set of input variables that can derive the direction of oil prices with high explanatory power post 2008 financial crisis. The effect of speculations and reserves together with EPPI is broken completely due to financial crisis. It shows that the influential mechanism of various factors on oil prices changed due to happening of geopolitical and economic events.

- **Superiority of proposed algorithms**
  Experiments showed that $I^2MI^2$ algorithm quickly identifies most relevant and non-redundant set of features. It has provided the minimal representative set of features which are more accurate in predicting oil prices as compared to other competing feature selection methods.

- **Number of features to be extracted**
  Without perturbing about the number of features to be extracted, on the natural domain, $MI^3$ and $I^2MI^2$ algorithm eliminated more than $\frac{1}{4}$ and $\frac{1}{3}$ of the features. The features thus selected through more refined version of $MI^3$ algorithm i.e. $I^2MI^2$ are 100% relevant and non-redundant features.
• **Application of proposed algorithm in varied disciplines**
  No single learning algorithm is superior to all others for all problems. But the proposed algorithm can provide the most relevant and non-redundant set of features for data mining problems. Practitioners can choose which algorithm to apply depending on their objective of the study. Armed with such insight, $I^2MI^2$ algorithm can enhance the performance of data mining problems, while at same time can achieve significant reduction in the number of features used in the study. $I^2MI^2$ algorithm can provide the minimal representative set of features for regression problems in business, biostatistics, applied energy and many more disciplines.

• **$I^2MI^2$ Algorithm is fully automatic algorithm**
  It doesn’t require user to specify any number of features to be extracted or to specify any threshold. It operated on original feature set and doesn’t incur the high computational cost associated with repeatedly invoking the learning algorithm.

• **Conditional Independence Assumption**
  Most of the feature selection methods assume that features are conditionally independent within the class. Due to existence of dependency within the features, these feature selection method could not perform well. This limitation is overcome in the proposed feature selection using the concept of interaction information. Conditional dependence is a measure of redundancy for complex real world problems. This research gap is taken care of in this study using proposed $I^2MI^2$ feature selection method.

### 6.3 Recommendations

The following recommendations are made to identify key drivers of oil prices.

• **Feature Selection for data mining problems**
  Forecasting oil price has never been an easy task, though it is important for so many economic policies. Using NYMEX future prices as stand-alone indicator for spot oil prices is not recommendable as there are high number of external factors influencing oil prices. Also, the input variables for any study cannot be selected based on judgemental criterion or trial and error method. It is a principal task to
identify key factors driving oil prices through feature selection algorithm before proceeding for model building and evaluation.

- **Influence of Geopolitical and Economic Events**
  The influence of input variables is assumed to be constantly driving oil prices in most studies. There is shift in influence of input variables driving oil prices subject to happening of geopolitical and economic events. These events are essential part of data analysis. Researchers are recommended to test for structural change in data due to happening of extreme events and further, proceed for model building and evaluation.

### 6.4 Limitations

- **Availability of data**
  Though Energy Information Administration (EIA) provides a comprehensive database for most of influential factors that can be incorporated in study but there are many more factors that are required to be accumulated for such important studies. There is no stability for the factors that can be considered for the study. Data is available for some factors in weekly or monthly term while some are available on quarterly or yearly basis. Also, researchers can predict oil prices more accurately if the forecasts of key factors is available over long term.

- **Selection of input variable for study**
  There is no exhaustive list of features that can be considered for study. Researchers have different view on finding relationship of oil price with supply-demand or with inventory, but these are not the only factors driving oil prices. The initial step of defining the dataset for the research ia a crucial step.

### 6.5 Contributions

This research leads to general contributions to the field of data mining and applied energy. The contributions are as follows:

- The study presented a new three stage $I^2MI^2$ algorithm for feature selection method, that performs very competitively as compared to
several state-of-the-art feature selection methods. The study presents both theoretical and empirical contributions. (Chapter-4)

- The study presents a new two stage \( MI^3 \) algorithm for oil price prediction that simultaneously improves the predictive performances for oil price predictions using significant input variables. (Chapter-4)

- The new proposed algorithms provides 100\% non-redundant and relevant features than previous feature selection methods for applications in varied disciplines. The explanatory power of key indicators influencing oil price market before and after financial crisis is presented. (Chapter-4).

- The study presents a new ensemble learning algorithm \((I^2MI^2 + GRNN)\) for prediction of oil prices with extensive empirical evaluation with EIA’s STEO econometric model (Chapter-4 & 5).

- A framework which can be used for predicting future value of oil prices depending upon movements in key factors driving the oil prices. (Chapter-5)

- The novel \( I^2MI^2 \) algorithm, which can be seen as a realization and an application of the proposed framework. Our experiments on real world problems show that the proposed algorithm performs very competitively as compared to other competitive ensemble models, and that it can provide optimized performance for real world complex problems. (Chapter-5)

6.6 Future Scope of the Study

The direction for future researchers are as follows:

- Detailed research can be carried out in subsequent studies by other scholars to quantify each factor for deriving directions of oil prices. Once these factors are quantified through separate research, the relevant and non-redundant features can be selected using proposed \( I^2MI^2 \) algorithm. Currently, this thesis has arrived at the fundamental stage of providing relevant and non-redundant features for any dataset.
• The study has been carried out to provide an insight into two major concerns: explanatory power of factors for oil price trend and their contribution to oil price prediction. The study can be expanded to provide information regarding transmission mechanism that follows between oil prices and factors.

• Future researchers can use number of other artificial intelligent forecasting engines with the proposed feature selection method to achieve high prediction performance for oil prices.

6.7 Concluding Remarks

The study has identified key factors influencing the direction of oil prices. China consumption and its reserves emerged as influential factors driving oil prices post 2008 financial crisis. The recent change in impact of Non-OECD consumption is highlighted in influencing oil prices as compared to OECD Consumption. OPEC Supply is dominating the fluctuations in oil prices due to sudden change in production targets or policies. With recent increase in Non-OPEC production, the influence of OPEC as the most influential factor driving oil prices is diminishing. NYMEX future price is not a stand-alone instrument for predicting spot oil prices but there are high number of external factors that are required to be identified. Since oil is traded in global market and most of the trade has operated and continue to operate in dollars, U.S Dollar Index remains an influential factor driving oil prices. Speculation and reserves played an important role in driving oil prices while CPI and EPPI have largest contribution as key drivers of oil prices after crisis.

The study showed the superiority of $I^2MI^2$ algorithm in comparison to other feature selection methods. Certain recommendations regarding the importance of each step in data mining process is highlighted in this chapter. Researchers are recommended to test for structural change in data due to happening of geopolitical and economic events. The contribution of the study in field of data mining and applied energy is presented together with future scope of the study.