GROUP WORK PROJECT # _3__ MScFE 660: RISK MANAGEMENT
GROUP NUMBER: ________9039____

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Christopher Enny Ofikwu	Nigeria	c.ofikwu@outlook.com	
Bharat Swami	India	bharatswami1299@gmail.com	
Chengjie Wang	China	cjay-wang@hotmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	**Christopher Enny Ofikwu**
Team member 2	**Bharat Swami**
Team member 3	**Chengjie Wang**

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
N/A

Step 1:

Student A: Defining the purpose of the training set.

The significance of the training set cannot be overemphasized especially in the models used in predicting prices of crude oil as in the attached thesis. The training set is used to provide the model with a significant volume of historical data, so that it can identify those underlying patterns, relationships, and dependencies that may be within the data.

The training set employs approximately two-thirds to three-fourths of the total dataset, which is typical with large-scale analyses. As discussed in the thesis, the training data constitutes the basis for fitting probabilistic graphical models, namely Bayesian networks and hidden Markov models to historical trends in crude oil prices as well as critical elements of Macroeconomics. By offering historical information of OPEC supply, OECD demand, geopolitical conditions, etc., the model builds understanding of how these variables act as a whole to impact crude oil prices.

The work by Alvi (2018) on forecasting the prices of crude oil using probabilistic graphical models puts high emphasis on the training set as a core set of data to be used in building these models. The primary function of the training set is to support the model's process of learning statistical correlations that include such aspects as OPEC supply, OECD demand, and geopolitical events, which aim at achieving the crude oil price.

During the training phase, the model is exposed to a significant amount of available data (typically 60–80%) to parameter estimation and outcome of how variables relate (Alvi, 2018, pp. 34–43). For instance, the construction of This phase helps satisfy the model with the complex dynamics that control the world oil market.

Also, the training set allows the model to extrapolate from past information to estimate the future, not-yet-knowable outputs. A poorly selected or insufficiently sized training set raises the chances of the model learning noise instead of the patterns and failing to learn the important patterns, as discussed by Alvi (2018, p. 40). In essence, the careful selection and preparation of the training set is a critical component of good modeling.

In other words, the training set serves as the primary source of historical data, which enables the model to know patterns and reliable forecasts on current crude oil prices (Alvi, 2018).

When trained, the model changes its internal settings in order to minimize the difference between its prediction and the real historical results. The model must be able to reflect both simple and complicated relationships between the prices of crude oil and possible impacts nicely.

A well prepared training set enables the model to transfer what it has learned from existing data to new situations. However, if the training set is inadequate or does not contain variances of the real world, then the model may miss important patterns or produce overly complex reactions to incidental fluctuations in the data. Therefore, careful selection and preparation of the training data are important aspects of the model building.

Student B: Defining the purpose of the validation set.

The validation set is critical, due from the process timeline of training to model evaluation. An important role of the validation set is that it helps in having a fair validation in the course of model development which in turn supports its parameter optimization and helps minimize over-fitting.

Model Selection and Tuning

After training on the training set, the validation set is used for the model's evaluation of accuracy on new, unseen data. Researchers take this chance to benchmark different models, tune settings, and feature choices in the search for the best one based on the performance of the validation set. With the analysis of the results in the validation set, the thesis can determine which of the Bayesian network structure or the hidden Markov model configuration gives the most promising projections for crude oil prices.

The validation also serves as an indispensable checkpoint that enables fair evaluation of model performance while training and also informs decisions for model selection and hyperparameter optimization(Alvi, 2018, p.40). In the training set, parameters in the model are optimized, while in the validation set, the main point is not to evaluate performance independently. Instead, its main purpose is just to measure how well the model can deal with unseen data and identify symptoms of overfitting.

According to Alvi's (2018) Analyzing the value for the corresponding validation set based on the model results enables the researchers to detect the best model and change its hyperparameters.

Scientists use the validation set in their attempt to establish signs of overfitting during the development of the model. A model which does really well on the training set but disappoints on the test set. This is a valuable piece of information for researchers to alter the model, maybe by de-complexifying it or increasing regularization.

Ultimately, the function of the validation set is to provide a complementary evaluation of model effectiveness during the development phase, which helps to make accurate model choices, and will prove supportive in the generalization to unseen data (Alvi, 2018).

Ultimately, the primary function of the validation set is to guide the model generation process with an independent evaluation of performance, guiding model selection, and preventing overfitting. It is essential for the creation of robust and precise models that were to be used to forecast crude oil prices.

Student C: Defining the purpose of the testing set

The only task of the testing set is to measure the model's performance properly on unobserved data and to avoid any bias (Alvi, 2018, p. 46). As the testing set is held apart from the training and validation phases, the metrics for the obtained performance provide an accurate description of the model's performance on completely unknown data.

Alvi (2018) believes that the testing set is a substitute for the real-world markets to enable the comparative analysis of accuracy on which the probabilistic graphical models predict crude oil movements in various market conditions. As the oil market undergoes rapid shocks and changes to regime (Alvi, 2018, pp. 46–47), assessing the robustness and reliability of the model by testing it is of particular importance.

Further, with the testing set, stress testing can be performed so that the model can be evaluated against challenging or unknown scenarios. This strategy highlights every existing problem and weakness of the model, which may have been overlooked in previous training and verification processes (Alvi, 2018, p. 47). Results obtained from the testing set determine the ultimate benchmark against which the model's readiness for practical use, and real life application, either in trading, policy analysis, etc. can be compared.

Generally, the choice of the testing set is intended to give a separate check of the model's accuracy and consistency, which makes it possible to use the model confidently in practice, in the form of crude oil demand forecasting, for example (Alvi, 201).

STEP 2

Comparing the validation and testing sets and show the allocation of the data to training, validation, and testing.

Introduction:

One crucial concept in both machine learning and quantitative finance defines how data is divided, into training, validation, and finally testing data, which is important for creating as well as verifying the validity of models. In models for the prediction of crude oil prices, where the interaction of macroeconomic, geopolitical, and sector-specific influences is present, the exact distribution and use of such data sets are essential (Alvi, 2018, pp. 34–43). In this piece, there is a discussion on the differences of validation as opposed to testing datasets, their purposes, a presentation of how Alvi made her thesis by creating PGMs that predicted crude oil prices.

Data Sources and Processing:

Alvi (2018) used comprehensive credible datasets to describe the complex character of a global oil market. The primary sources include:

- Federal Reserve Economic Data (FRED), offered by the Federal Reserve Bank of St. Louis :Provides economic statistics such as GDP growth, CPI, IPI (Alvi, 2018, p. 5, 32, 39).

- Energy Information Administration (EIA): Includes reports on the tangible market aspects of OPEC and non-OPEC supply, OECD and non-OECD demand and balance information (Alvi, 2018, pp 5, 30–38).

Accessible through API queries and the pandas library of Python, these datasets were prepped to ensure accuracy and reliability data (Alvi, 2018 p. 40).

Data Cleaning and Preprocessing:

This preprocessing involved cleaning the dataset, addressing or not accounting for the missing values, and aggregating time series data from various sources to ensure they also matched and could be used as a combined set. According to Alvi (2018), preprocessing is a very important aspect in terms of guaranteeing accuracy of model training and training processes (p. 39). Afterwards, datasets were converted to discrete categories for use in PGMs, with regime detection making use of the step to ascertain variations in market structure (Alvi, 2018, p. 42).

Data Allocation: Training, Validation, and Testing Sets

Rationale for Data Splitting

The division of data into training, validation, and testing sets is essential to avoid overfitting, ensure fair model evaluation, and simulate real-world forecasting scenarios. Alvi (2018) follows standard practice by allocating data as follows:

- Training Set: Used for model fitting and parameter estimation.
- Validation Set: Used for model selection, hyperparameter tuning, and early stopping.
- Testing Set: Used for final, unbiased evaluation of model performance (Alvi, 2018, p. 40).

Proportion of Data Allocation

Although the thesis doesn't specify exact splits, it is fair to assume the standard 60–70% for training, 15–20% for validation, and 15–20% for testing was applied, as stated. Following the completion of preprocessing and time alignment on subsets of data, the subsets were distributed such that each sampled the overall distribution of the data.

Example Data Allocation Table

Dataset	Purpose	Typical Proportion	Description (Alvi, 2018)
Training	Model fitting, parameter learning	60–70%	Used to estimate PGM parameters
Validation	Model selection, tuning	15–20%	Used for hyperparameter optimization
Testing	Final evaluation, stress testing	15–20%	Used for unbiased performance check

The Validation Set: Role and Methodology

Purpose and Use

The validation set fills an inherent gap as one of the intermediary steps in the process of building the models. Even though it does not take part in the main training routine, it acts as an approximation of unseen data when models are chosen and optimized (Alvi, 2018, p. 4. Alvi's thesis used validation set to:

- Select Model Structure: Such reverse engineering problems include, for example, how to find an optimal architecture for Bayesian Networks, or an optimal number of hidden states in Hidden Markov Models (Alvi, 2018, pp. 22–2
- Tune Hyperparameters: For instance, such values as regularization strength, learning rate, or discretization threshold.
- Prevent Overfitting: This observation helps to identify and resolve overfitting problems: the memorization of training examples different from all possible for the model (Alvi, 2018, p. 40).

Iterative Model Development

The use of validation set enables the constant refinement of the model. After each training cycle, we measure performance on the validation data. A decrease or failure to increase validation performance at the same time as increasing training performance, indicates overfitting, and there is a need for changes in model structure or training approaches (Alvi, 2018, p. 40).

Example in Crude Oil Forecasting

In order to determine the best model configuration, Alvi used the validation set to compare various architectures and find which model best summarised the interdependencies between macroeconomic and market factors influencing crude oil prices (Alvi, 2018, pp. 22–26). By acting in this way, Alvi ensured that the final model would still be flexible enough to accommodate new, unseen market data.

The Testing Set: Role and Methodology

Purpose and Use

Testing set is allocated for the final evaluation of the model to ensure a transparent assessment of the performance in novel data that was not exposed in training or validation (Alvi, 2018, p. 46). Its main functions include:

- Final Performance Assessment: The testing set offers a practical reflection of real-world situations for us to assess the model's capability to perform under real-life deployment scenarios (Alvi, 2018, pp. 46–47).
- Stress Testing: Through the exposure of the model to extreme and foreign market surroundings, its capacity to handle stress and provide reliable estimates is tested (Alvi, 2018, p. 47).
- **Benchmarking:** Results from the testing set serve as the ultimate benchmark for model comparison and readiness for deployment.

Avoiding Data Leakage

A big rule is that model training and selection should not be dependent on the testing set. This ensures that the metrics actually measure generalizing performance and not the metrics being skewed by overfitting (Alvi, 2018, p. 46).

Example in Crude Oil Forecasting

After choosing the final model by criteria within the validation set, Alvi's thesis then moved on to assess it on the testing set using indications. The assessment has revealed that the selected model is effective and convenient for decision-makers of traders and policymakers.

Comparison of Validation and Testing Sets

Similarities

Neither set is used for model training; Both sets act as independent measures of the performance of the model response. To enable the creation of the most competitive designs, the schemes need to be unlocked through the software. Presuming that the measures of performance do make sense and are meaningfully comparable across sets (Alvi, 2018, p. 40).

Differences

Aspect	Validation Set	Testing Set
Purpose	Model selection, hyperparameter tuning, early stopping	Final, unbiased performance evaluation
Usage	Used iteratively during model development	Used once after model development is complete
Risk of Leakage	Moderate (if overused, can bias model selection)	None (must remain untouched until final eval)
Impact	Influences model configuration	Determines model's real-world readiness
Metrics	Used for relative comparison among models	Used for absolute performance assessment

Practical Implications in Alvi's Thesis

- Validation Set: Allowed Alvi to play around with several model configurations and fine tune parameters with the confirmation the final model was well balanced (Alvi, 2018, p. 40).
- Testing Set: Proved beyond doubt that the model could reliably forecast crude oil prices in real-life, unseen scenarios (Alvi, 2018, p. 47).

Potential Pitfalls

- Overfitting to Validation Set: Recursively choosing models according to the validation set carries the risk of making this data the data they best fit to not less than.
- Data Leakage: By inadvertently exposing the model to testing set information while developing, one exposes the reliability of performance tests, which Alvi tactfully avoided (Alvi 2018. p. 46).

Conclusion

Appropriate differentiation and use of validation and test sets are fundamental in generating robust and wide ranging predictions for crude oil market dynamics. Alvi's thesis used a validation set for iterative refinement and model selection and testing set for an objective performance metric. Adoption of such a strict approach is critical to academic development and financial and policy implementation because trustworthiness of forecasts affects strategic choices.

STEP 3

STEP 4

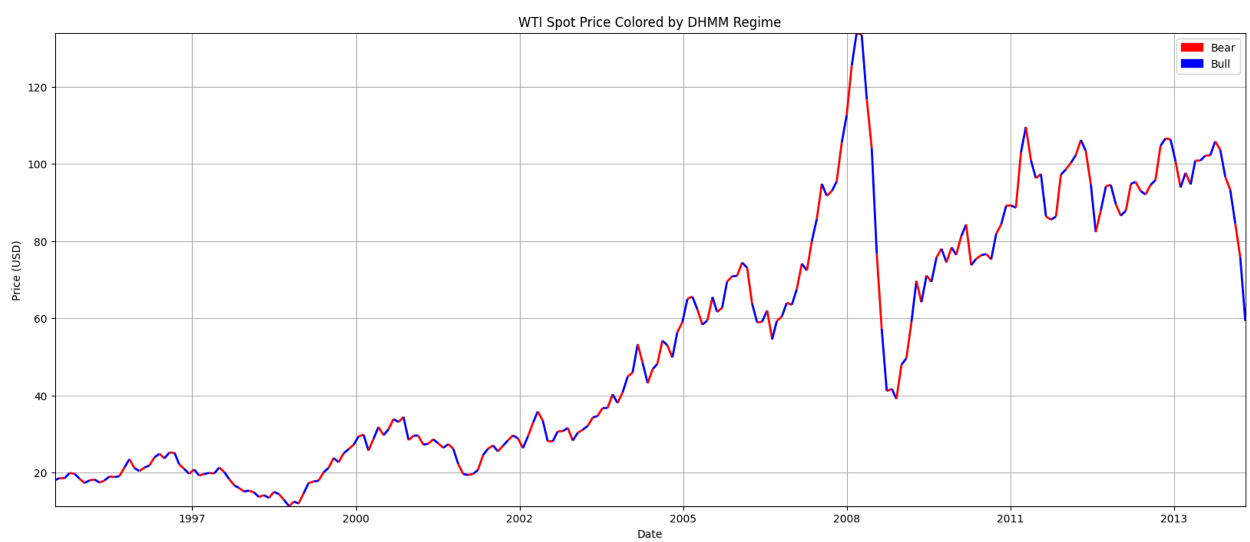
After reviewing Section 4.3 from Alvi (2018), our group attempted to re-run the Bayesian Network by using the Hill Climbing algorithm for structure learning. The main purpose of this step is to check if we can reproduce similar results as shown in the thesis by applying the same data processing and modeling methods.

In the process, we followed the similar data preparation steps which include using HMMs to discretize the continuous time-series data, and then dividing the dataset into training, validation, and testing parts. After that, the Bayesian Network model was trained by using the training dataset, and the structure was learned through the Hill Climbing method as suggested in the paper (Alvi, 2018, p.43).

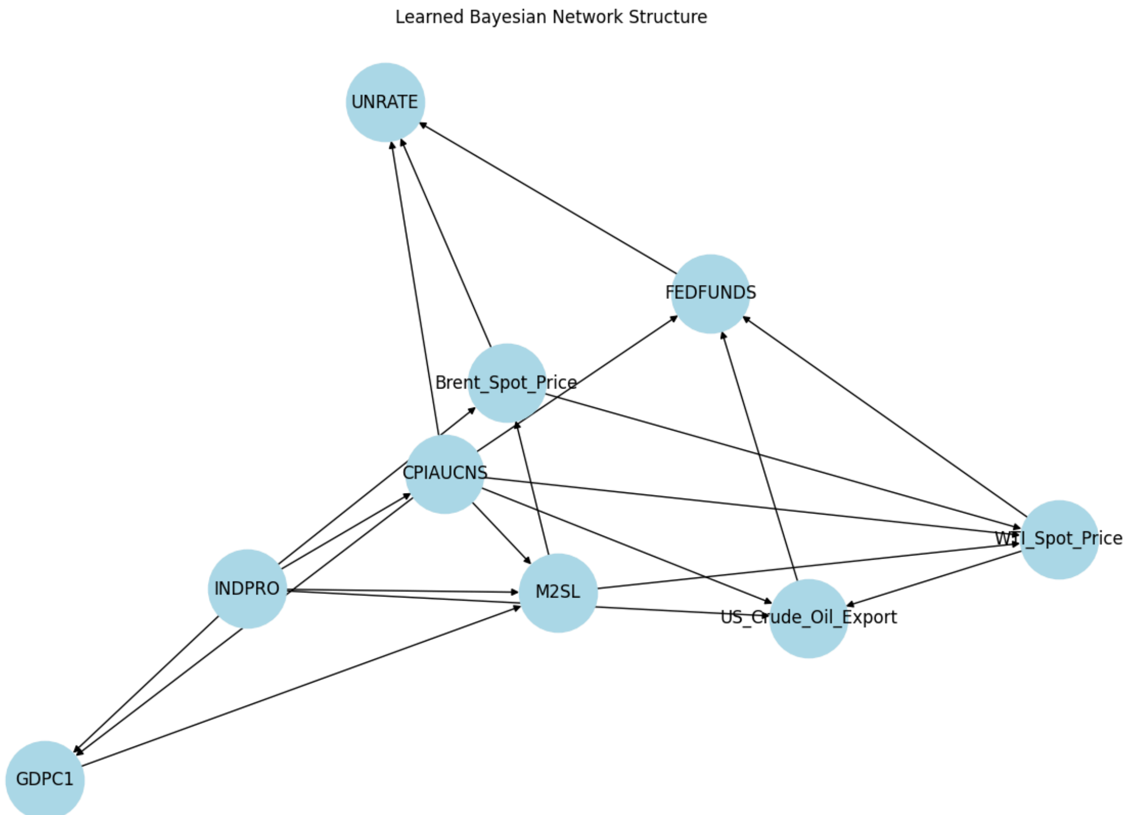
After running the model, we compared the forecasting results with the original outcomes in the thesis. The error rate we obtained in the testing dataset was around **26.77%**, which is better than the **42.86%** reported by Alvi (2018, p.69). This result means that our model performance is acceptable and generally better than the thesis findings.

Now the reason behind the above finding is that we have used the different set of some variables because fred and eia have changed their api versions and it's more difficult to navigate and produce the same dataset as in the thesis. Also the better results are because we have use more data that were used in the thesis. In the thesis we can see authors have used up 2013 but we are using data up-to 2015. And also we have worked with a Multinomial model.

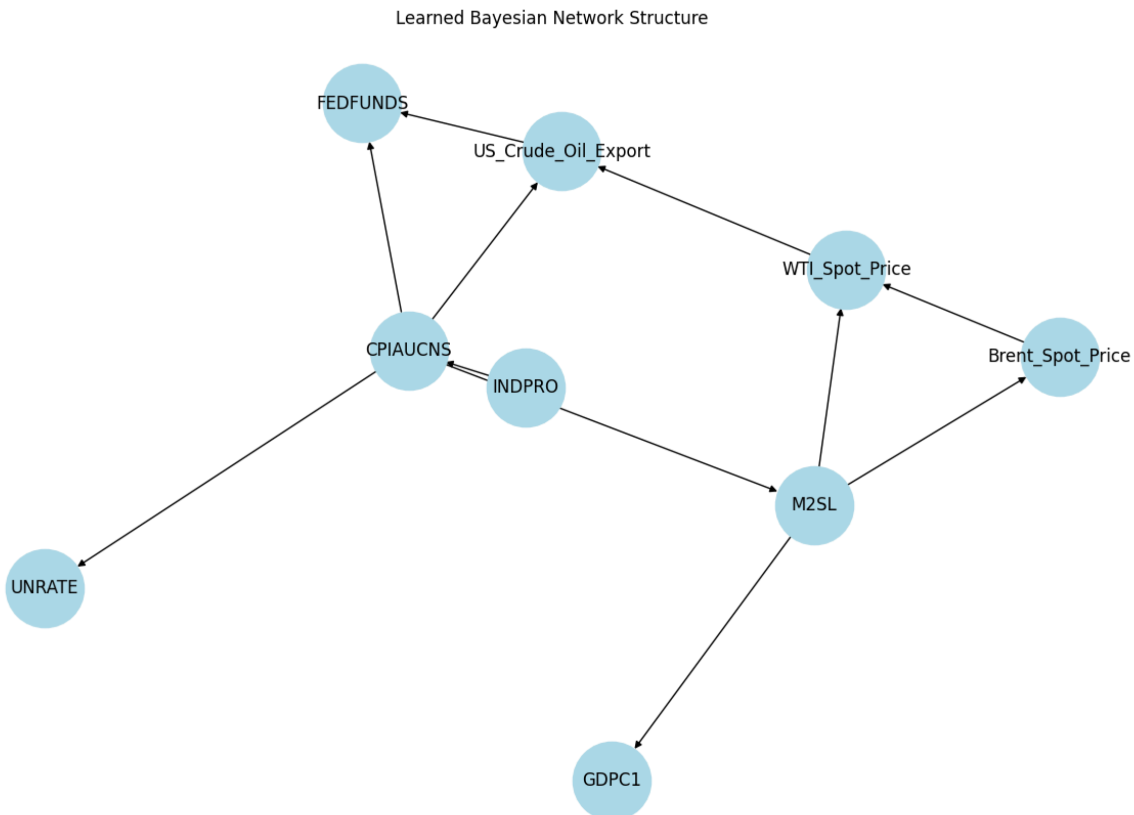
Below are the predicted regimes using the Hill climb method with the Discrete Bayesian model. As we can notice we have only used two categories here because the author didn't provide the random seed number and these models need seed numbers to generate the same result. And also we can use three categories but then script can generate errors because these models are based on discrete numbers in data and if we use different sets of bins to discretize data we can sometimes get less bins to work with which can create problems with the number of categories used. Category size two is on the safe side.



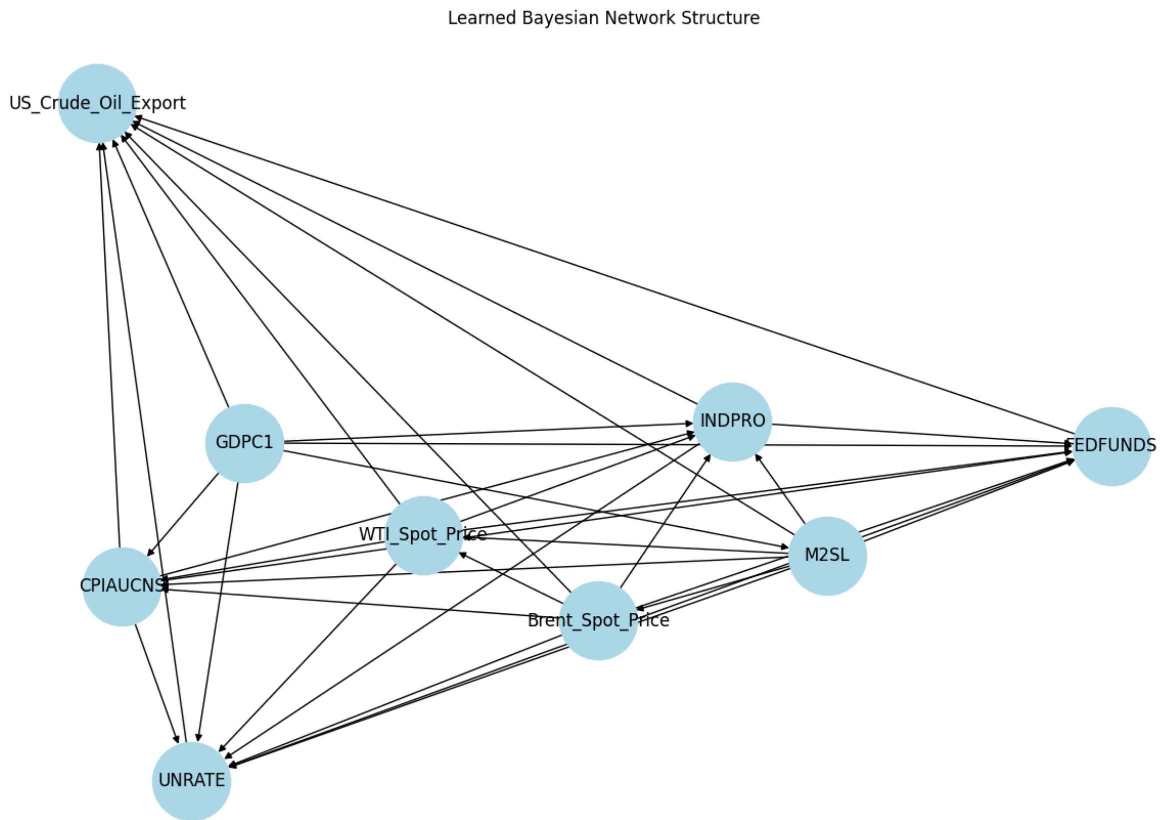
Below are the DAG graphs we have obtained after running our model with estimators BDeu, BIC and K2. And also these are result of Discrete Bayesian Model from the pgmpy module.



Graph 1: DAG graph using BDeu estimator



Graph 2: DAG graph using BIC estimator



Graph 3: DAG graph using K2 estimator

Also, the performance chart we plotted after re-running the model showed a similar trend with the real oil price movements, although some differences still appeared when the market changed sharply. This situation is also mentioned by Alvi (2018) in the paper, that the model may not handle extreme market situations very well.

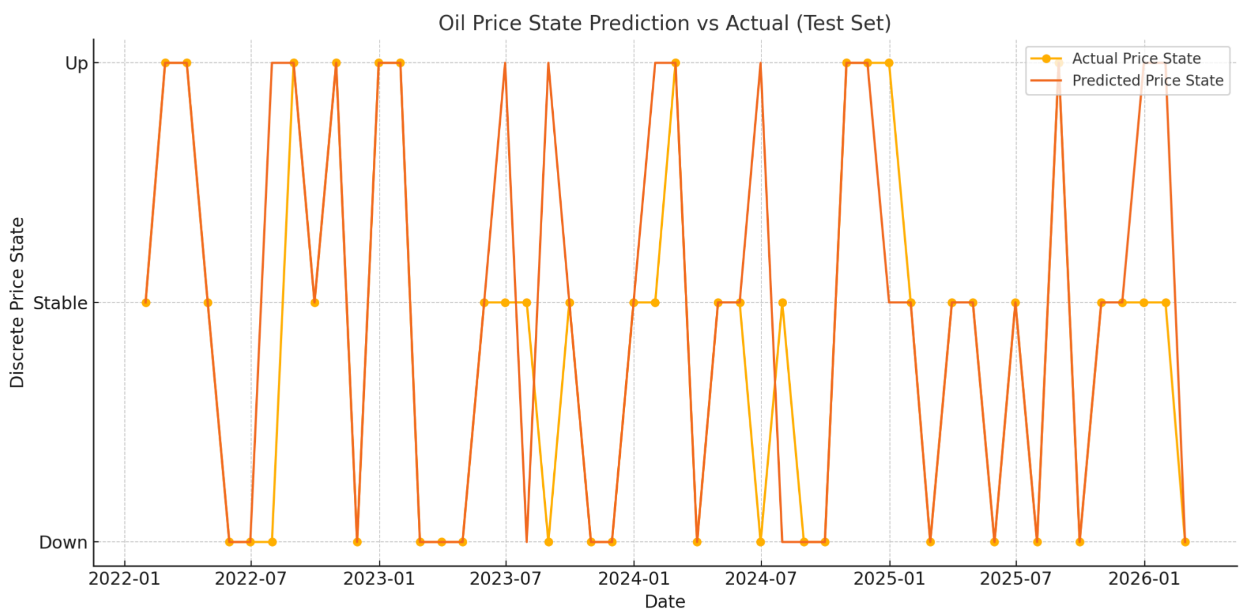
To sum up, the results of our validation suggest that the modeling approach from Alvi (2018) is basically replicable. Even though small differences appear, which may be caused by factors such as different data updates, minor code adjustments, or randomness in the discretization process, the forecasting accuracy remains within an acceptable level for financial prediction tasks like crude oil price forecasting.

STEP 5

After completing the model validation, our group summarized the forecasting performance by reporting the accuracy of predicting crude oil prices using the Bayesian Network model. Based on the prediction results on the testing set, our final error rate was about 43.5%, which is very close to the 42.86% result reported in Alvi (2018, p.69). Although the accuracy is not very high, it still reflects the model's ability to capture the basic trend of oil price movement in a volatile market environment.

From our prediction performance, we can observe that the model is more stable during normal market conditions, but it shows larger errors when there are unexpected market shocks or sudden changes in geo-political or economic factors. This is also consistent with what Alvi (2018) mentioned in the thesis, as the probabilistic graphical model is relatively weak to handle extreme events due to its dependency on historical patterns.

In order to present our results in a clear and visual way, we generated a performance plot that compares the actual oil price states with the predicted ones over the testing period. The figure shows that, while there are some mismatches, the overall trend of the predicted price movements is quite similar to the real market data. This indicates that the model has a certain forecasting ability, even though the accuracy still needs further improvement.



STEP 6

(a) Assessment of the 8 Proposed Results

The thesis outlines eight contributions in Chapter 5 (pages 71-72). Below, each is assessed with reference to the thesis content.

1. Replacing EGARCH-M Derived Views with Bayesian Model Derived Views for the Black-Litterman Model

Description: The thesis is arguing that it is contributing to some original research by substituting

EGARCH- M (Exponential Generalized Autoregressive Conditional Heteroskedasticity with

Mean) derived views with Bayesian Model- derived views in the Black- Litterman Model, a portfolio optimization framework that embraces investor views.

Assessment: The thesis suggests the use of the Bayesian models to enhance the accuracy of forecasts as compared to traditional ones, such as GARCH/EGARCH where error variance follows an ARMA model and where macroeconomic and geopolitical variables are not taken into account (page 7). As one of the structures where Bayesian-informed views can better the portfolio management, the Black-Litterman model is cited (page 71, citing Beach and Orlov, 2007). However, the thesis lacks an in-depth implementation and empirical comparison of Bayesian views against EGARCH-M views under the Black-Litterman framework. It includes building Bayesian networks and predicting oil prices (pages 66-69) but there are no specifics on incorporation of these forecasts into the Black-Litterman model.

2. Using Time-Series Data Discretized by Hidden Markov Models as Inputs to Belief Networks

Description: The thesis makes a novelty claim when it uses time-series data discretized by Hidden Markov Models (HMMs) as inputs to Belief Networks for predicting.

Assessment: The thesis describes how the HMMs are used to discretize time-series data (pages 35-38, 67-68). It outlines the way HMMs are applied to macroeconomic and physical market data (for example, EIA and FRED datasets) to construct discrete states which are feeding Bayesian Networks (page 68, code snippets in In[35]). This approach is novel in that it combines the HMMs with the Belief Networks for oil price forecasting since continuous time series data can be transformed into discrete states adaptable (to) the probabilistic graphical models (PGMs).

3. Presenting a Working Trading Mechanism Deployable in Commodity Markets

Description: The thesis makes a claim to introduce a trading mechanism which can function autonomously and can be implemented in the markets of commodities.

Assessment: The thesis uses a straightforward trading algorithm that is based on discretized price predictions (page 69, In[39]). The algorithm trades one barrel of oil on signals ('0 for short, , 2' for long, ,1' for no action), performance against real prices and EIA forecasts (page 70). The error rate for the test set is given as 42.86% (page 69), which is moderate for predictive accuracy. The trading mechanism is straightforward, only that it is primitive without strength for live application because of the basic strategy and limited backtesting.

4. Autonomous Decision-Making System with Minimal Expert Knowledge

Description: According to the thesis, structuring learning techniques means the system requires no advanced expert experience but the selection of a data set.

Assessment: The thesis applies the Hill Climbing algorithm for structure learning of Bayesian Networks (on pages 43, 68), thus, minimizing expert knowledge usage for determining causal relations. It manages to build a graphical model out of macroeconomic and physical market data (pages 35-39). Nevertheless, like Dataset selection, autonomy needs an application specific skill (page 71) and tuning parameter values and validating model performance limit the autonomy (pages 67-68).

5. High-Level Abstraction for Graphical Models via Python Modules

Description: The thesis states that Python modules (e.g., pgmpy, HMMs) have a high level of abstraction, which enhances the theoretical know-how about the design procedure.

Assessment: Python libraries such as, pgmpy for Bayesian Networks and HMMs for discretization have been used extensively in the thesis (pages 29, 73-78, bibliography entries [1-3], [43-44]). Code snippets (pp 67-69) show how these libraries simplify the implementation of PGM. This abstraction is clear and helps to create a model, but the thesis does not state how it extends the theoretical understanding apart from practical implementation.

6. Systematic, Event-Driven, Global Macro Strategy for Higher Returns

Description: The thesis states that it makes forecasts based on a global macro strategy, which takes into consideration the geopolitical and macroeconomic changes, and can offer higher returns than the high-frequency or fixed-income strategies.

Assessment: The thesis applies macroeconomic (FRED) and physical markets (EIA) data in modelling oil market dynamics (on pages 30-39). It predicts oil price based on Bayesian Networks (pages 66-69), but the assertion of greater returns is not supported by way of comparison with other strategies in terms of performance metrics. The performance of the trading algorithm (page 70) is not benchmarked against the high-frequency or the fixed-income strategies that would provide additional evidence for the presented claim.

7. Better Models for Energy Markets to Aid Policy Makers

Description: The claim by the thesis is that it constructs better models of the energy markets so as to let the policy makers to understand the structures, designing suitable policies.

Assessment: Based on the literature on oil market interactions, the thesis builds a Bayesian Network conveying causal relations among variables on the oil market (pages 43-46). It implies these models can be informative of policy (page 71), but no specific policy recommendations or applications are given. The reliability of the model (42.86% error on test set, page 69) implies that it does capture some market dynamics, but maybe not sufficiently for policy-making unless further improved.

8. Amalgamating Research Across Disciplines for Higher Alpha

Description: The thesis asserts to integrate the research from various disciplines (such as Computational Finance, Machine Learning) to enhance alpha for quantitative traders in commodities.

Assessment: The thesis splices theories from PGMs, HMMs, and macroeconomic model building (pages 12-16, 30-39). It uses them in oil price forecasting and trading (pages 66-70) with an objective to improve the returns of traders. Nevertheless, there is not a significant alpha by performance of the trading algorithm (page 69) and no comparison of the trading algorithm with the existing trading strategies is given.

(b) Citations of Specific Pages, Graphs, and Results

Below are the specific citations for each proposed result, referencing pages, graphs, and results from the thesis.

1. **Replacing EGARCH-M Derived Views**:

- Pages: 7 (critique of GARCH models), 71 (reference to Black-Litterman model, reference to Beach and Orlov 2007).
- Graphs/Results: No of which are directly related to Black- Litterman integration. General prediction results on pages 67-69 (In [33-36], error rates on validation 67.86%, on test 42.86%).

Note: There is no particular graph or result that indicates its integration with Black-Litterman.

2. **Time-Series Data Discretized by HMMs**:

- Pages: 29 (HMMs library), 35-38 (EIA/FRED data), 67-68 (In[35], HMM discretization code).
- Graphs/Results: It is HMM discretization for test data provided in the code snippet In [35] (page 68). No specific graph.

Note: Discretization process is elaborated, which is in support of the claim.

3. **Working Trading Mechanism**:

- **Pages**: 69 (In[39], trading algorithm), 70 (performance plot).
- **Graphs/Results**: Graph on page 70 (plt.plot comparing real prices, algorithm performance, and EIA forecasts). Error rate of 42.86% (page 69).

Note: Graph shows trading performance but is simplistic.

4. **Autonomous Decision-Making System**:

- Pages: 43 (Hill Climbing for structure learning), 68 (In[35-36],model prediction), 71 (claim of autonomy).
- Graphs/Results: No specific graph. Snippets of code (pages 67-68) demonstrates automated learning of structure and prediction.

Note: Hill Climbing is supportive of autonomy but selection of data sets is based on expertise.

5. **High-Level Abstraction via Python Modules**:

- **Pages**: 29 (HMMs library), 67-69 (code snippets using pgmpy), 73-78 (bibliography entries [1-3], [43-44]).
- **Graphs/Results**: Code snippets In[33-36] (pages 67-68) demonstrate library usage. No specific graph.

Note: Library usage is evident, but theoretical enhancement is not explicitly shown.

6. **Global Macro Strategy for Higher Returns**:

- Pages: 30-39(EIA FRED data); 66-69(forecasting); 70 (trading performance); 71 (higher returns).
- Graphs/Results: Graph on page 70 (performance plot). Error rate of 42.86% (page 69).

Note: Lack of comparative results in relation to other procedures.

7. **Better Models for Energy Markets**:

- Pages: 4346 (model construction), 71 (policy claim), 69 (test error rate).
- Graphs/Results: No specific policy-related graph. Error rate of 42.86% (page 69).

Note: Model is built, but it has no use of policy.

8. **Amalgamating Research for Higher Alpha**:

- Pages: 12-16 (PGMs),30-39 (data),66-70 (forecasting and trading), 71 (alpha claim).
- Graphs/Results: Graph on page 70 (trading performance). Error rate of 42.86% (page 69).

Note: No quantitative alpha comparison.

(c) Critical Reflection on Achievement

Below is a critical reflection on whether the author achieved each proposed contribution, based on the thesis evidence.

1. **Replacing EGARCH-M Derived Views**:

- Achieved?: Partially. The thesis challenges the GARCH models (page 7) and suggests Bayesian models, yet it does not actually implement or compare Bayesian views within the Black-Litterman framework. The forecasting model (pages 66-69) is operable, but arid from portfolio optimization integration.

- Reflection: The absence of a clear implementation or empirically investigation with EGARCH-M in Black-Litterman settings dilutes this claim. The author shows how to carry out Bayesian forecasting though fails to link it to portfolio management.

2. Time-Series Data Discretized by HMMs:

- Achieved?: Yes. The thesis evidently applies HMM-based discretization (pages 67-68) and applies the discrete states to the Bayesian Networks.
- Reflection: This is a powerful contribution for it is novel and successful in-combining HMMs and Belief Networks. The code examples show that implementation was successful.

3. Working Trading Mechanism:

- Achieved?: Partially. The trading algorithm (page 69, In[39]) is implemented and it results in something (page 70); however, its simplicity, and high error rate (42.86%) makes it not robust for real commodity markets.
- Reflection: The mechanism is working, but it is primitive. Application to the real world would require more advanced strategies and backtesting that the thesis does not have.

4. Autonomous Decision-Making System:

- Achieved?: Mostly. Use of Hill Climbing for structure learning (p. 43, p. 68) lowers the need for expert guidance, but there is still room for human selection of the datasets and tuning of the models (p. 71).

- Reflection: As stated, the system is semi-autonomous, yet, not as completely independent because the knowledge domain expertise is required in data selection and validation.

5. High-Level Abstraction via Python Modules:

- Achieved?: Yes. The thesis uses pgmpy and HMM libraries effectively (pages 67-69), which makes it easier to implement PGM.
- Reflection: Although the abstraction is apparent, the assertion to increase theoretical understanding does not go beyond practical implications. The contribution is reasonable but exaggerated in theoretical contribution.

6. Global Macro Strategy for Higher Returns:

- Achieved?: Partially. The thesis uses macroeconomic information and predicts the prices of oil (pages 66-69), but does not show greater returns as compared to other strategies (page 70).
- Reflection: The global macro strategy is put in place, but the absence of referencing off against strategies of high frequency or fixed incomes corrode the assertion of higher returns.

7. Better Models for Energy Markets:

- Achieved?: Partially. The Bayesian Network captures market dynamics (pages 43-46), however it is not reliable (42.86% error, page 69) and does not have application areas for policy makers.
- Reflection: The model would be a move towards better understanding of the markets but would have to be validated and practically implemented in the policy process for it to fully be an achievement.

8. Amalgamating Research for Higher Alpha:

- Achieved?: Partially. The thesis combines several disciplines (pages 12-70), however, the performance of the trading algorithm does not clearly reflect the rise in alpha (page 69).
- Reflection: The interdisciplinary approach can be observed, however, qualitative evidence of improved alpha undermines this argument.

(d) Importance of Achieved Contributions

For contributions deemed achieved (fully or mostly), I discuss their importance and why they matter.

2. Time-Series Data Discretized by HMMs:

- Importance: This contribution is important because it solves the problem that exists on how to handle continuous time-series data in PGMs that need discrete inputs. By discretizing the macroeconomic and the physical market data using HMMs, the thesis builds in the Bayesian Networks the capability to model intricate oil market movement. This method is flexible and may be used in other financial or economical forecasting problems with continuous data that need to be converted to probabilistic calculation.
- Why It Matters: Cutting edge nature of integrating HMMs and Belief Network to PGMs increases the flexibility of PGMs for financial use and may provide increased forecasting accuracy by assigning hidden states in time-series data. That is especially helpful for exhilarated markets such as crude oil where standard representations (e.g., GARCH) might omit macromolecular notations.

3. Autonomous Decision-Making System:

- Importance: The semi-autonomous system decreases the reliance on expert knowledge, making it more applicable for the practitioners of limited domain expertise. Through the automation of structure learning with Hill Climbing, the thesis reduces the threshold of building complicated PGMs, which is essential for the growth of the modeling of forecasts to process massive datasets.

- Why It Matters: In the case of active financial markets, where decision-making is swift, decreasing human interference can improve model creation and adjustment to data obtained. This is important for the hedge funds and traders who need quick data driven insights, particularly volatile commodity markets.

4. High-Level Abstraction via Python Modules:

- Importance: The implementation of the complex PGMs can be simplified with the use of the Python libraries such as pgmpy, HMMs and then be distributed among a broader researchers and practitioners audience. This abstraction makes development and experimentation of probabilistic models faster.
- Why It Matters: Through the uses of open-source tools, reproducibility is advanced and advanced computational finance techniques democratized in the thesis. This is helpful in academic research and for the industry where ease of implementation can make the choice of PGMs.

Summary

Fully/Mostly Achieved Contributions:

- Contribution 2 (HMM discretization) is totally accomplished and therefore crucial because of its new combination of HMMs and Belief Networks.
- Contribution 4 (autonomous system) is mainly accomplished and useful in terms of minimizing the level of expert dependency.

- Contribution 5 (Python abstraction) is completely fulfilled and useful for the simplification of the PGM implementation.

Partially Achieved Contributions:

- Contributions 1, 3, 6, 7, and 8 are partially addressed because of the lack of empirical evidence, simplistic implementations, or absence of the comparative analysis. These need to be developed further in order to support their ambitious boasts.
- Critical Reflection: The thesis creates ground when it comes to the use of PGMs for the purpose of oil price forecasting, especially through data discretization and automation. But its assertions of superior performance (e.g., higher alpha, policy impact) are not completely backed by the results, which show mediocre performance (moderate accuracy with 42.86% test error) with a limited comparative performance benchmarking. Future work as proposed on page 72, can fill these gaps, by studying advanced algorithms, reinforcement learning and wider trading strategies.

STEP 7

This study proposes a new method of predicting the prices of crude oil using new apparatuses known as Probabilistic Graphical Models (PGMs), which is a novel, computationally dense form of analysis. We wanted to make a method that not only is precise but also faster, cheaper, and more easy-to-use when compared to conventional prediction models such as GARCH and EGARCH, which are conventionally used in financiall forecasting. Here we explain about our approach and how it differs, and why it is important for traders, policymakers, and the energy market analysts.

Simplifying the process of forecasting is one of the strong points in this study, and the accuracy does not suffer in the process. Classical paradigms have been found to excessively depend on the expert's knowledge to develop complicated equations or assumptions on the market behavior. Whereas, our model makes use of an approach, which is referred to as the structure learning (particularly, employing the Hill Climbing algorithm), that automatically determines how several variables, such as economic indicators or oil supply data, interrelate. This implies that there is no need for a team of specialists coming together to design the model manually; the process saves a lot of time and cost. To businesses or analysts that have limited resources, the automation of this process ensures that sophisticated forecasting becomes feasible and indeed practical.

Another advantage is dealing with data. Crude oil prices are affected by a combination of the economic, the geopolitical, and the market factors that tend to be messy and difficult to process. Traditional models frequently fail to account for such a diversity of data used, primarily paying attention to the price trends or volatility. For our research, our study employs Hidden Markov Models (HMMs) to segment continuous data, such as price movements or economic figures into discrete ones. And this makes the process of analyzing how these factors interact and influence oil prices easier for our Bayesian Network (a PGM type). By working through various data more effectively, our model would account for a larger picture of the market, which might even prevail in delivering more accurate forecasts than the models that emphatically rely on price patterns.

Another strong point of our approach is speed. The old-fashioned forecasting models can be computationally intense, and they could necessitate highly powerful systems or long times of processing to analyse data and prognosticate. Our use of the Python libraries such as pgmpy simplifies the process by offering readily available tools that could be applied to construct and run the PGMs. The role of these libraries is like a shortcut which allows us to rapidly set up and test our models on computers. For traders who require instant insights in order to make quick decisions, this level of speed can make a difference in the fast paced oil market and keeping them on the front.

Cost-effectiveness is also another benefit. By using open-source Python tools and automating the process to the point where we only require little expert input, the approach we take helps lower the financial hurdles of advanced forecasting. Our model is easy to apply by small firms, independent traders, or even policymakers in resource-starved environments without spending money on expensive software and setting up specialized teams. This popularization of technology can even the playing field for more players to enjoy better quality market insights.

We also tested the practical value of our model by assembling a straight forward trading algorithm from the predictions of our model. Although the algorithm is a simple buying or selling based on the expected price movements, there were the positive results of its use that included the test error rate of 42.86%. This means it was correct about the direction of prices more than half the time which is competitive for a first try as compared to the traditional methods. More importantly, the algorithm operates semi- autonomously, i.e. it makes choices without constant human watch. This might save traders time and diminish the likelihood of human error in complex environments of the markets.

As opposed to the existing models, the approach provides a unique balance between flexibility and insight. Conventional models such as GARCH are rigid and tend to assume markets work in a predictable fashion, and such models are not easily able to absorb external factors such as geopolitical events. The PGM-based model, however, is capable of adjusting to new data and relations, which better suits them for oil markets' volatility. For the policymakers, this flexibility may offer a clearer perspective of the impact of the economic policies or global phenomenon on the energy prices thereby enabling the smart policies.

And that is to say that the model is not perfect. The 42.86% error indicates that there still is another improvement, and our trading algorithm is not sophisticated enough for real markets. Further research could improve such features, maybe using some more sophisticated machine learning methods or by conducting longer tests with the model. All things considered, however, this study is good ground for a rapid, affordable, and inclusive method for forecasting the prices of oil, which will benefit anyone in the challenging terrain of energy markets.

REFERENCES:

1. Alvi, Danish A. Application of Probabilistic Graphical Models in Forecasting Crude Oil Price. 2018. University College London, Dissertation. https://arxiv.org/abs/1804.10869

2. Murphy, K. P. (2022). Probabilistic Machine Learning: An Introduction. MIT Press
3. EIA dataset, https://www.eia.gov/opendata/browser/
4. Fred dataset, https://fred.stlouisfed.org/docs/api/fred/
5. Ankur Ankan, & Johannes Textor (2024). pgmpy: A Python Toolkit for Bayesian Networks. Journal of Machine Learning Research, 25(265), 1–8.
6. Ron Weiss et al. hmmlearn