Engineering Journal issn: 0125-8281

Article

High-Density Polyethylene Film Price Forecast in Southeast Asia Market with Deep Learning

Narongchon Amapat[1,a] **and Parames Chutima**[1,2,3,b,*]

1 Regional Centre for Manufacturing Systems Engineering, Chulalongkorn University, Thailand
2 Industrial Engineering, Engineering Faculty, Chulalongkorn University, Thailand
3 The Academy of Science, The Royal Society of Thailand, Thailand
E-mail: [a]narongchon.a@gmail.com, [b,*]parames.c@chula.ac.th (Corresponding Author)

Abstract. Neural networks (NN) have been used for over a decade to predict time series data, with various algorithms including linear and non-linear models. Forecasting and assessing polymer market prices are crucial for plastic resin producers due to the complexity and uncertainty of resource availability. This encompasses feedstock planning, raw material procurement, technological advancements for product transitions, sales planning, pricing strategies for commercialization, and investments driven by macroeconomic factors. Previous literature primarily utilized numerical data as input for deep learning models. This research contended that structured data by itself was inadequate for models to precisely predict outcomes in the volatile, uncertain, complex, and ambiguous (VUCA) environment. Three deep learning architectures, Long-Short Term Memory (LSTM), Encoder-Decoder, Temporal Convolutional Network and Recurrent Neuron Network (TCNRNN), were reviewed in this research to determine the most effective architecture for analysing structured data. Additionally, Natural Language Processing (NLP) was implemented in this research to gather market sentiment and enhance forecast accuracy. The study utilizes commodity market price announcements, economic indicators, and insight reports from reputable publishers. The study utilizes commodity market prices, economic indicators, and insightful reports. All information was obtained from a reputable publisher. The results were compared with the legacy model, which involved a human analyst and a linear regression model. Model performance was assessed using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and ANOVA. The linear regression forecast together with the human analyst model has an acceptable accuracy with a MAPE of 45.1%. Neural networks containing sentiment analyzers have been found to surpass the performance of human analysts and a linear regression model, with a MAPE of 17.1%.

Keywords: Inflation forecasting, neural networks, time series models, hybrid models, partner countries' inflation, forecast accuracy, deep learning, Bank of Thailand, economic indicators.

ENGINEERING JOURNAL Volume 28 Issue 12
Received 19 July 2024
Accepted 6 December 2024
Published 31 December 2024
Online at https://engj.org/
DOI:10.4186/ej.2024.28.12.79

DOI:10.4186/ej.2024.28.12.79

1. Introduction

Polyethylene (PE) is a thermoplastic material that is extensively manufactured and finds application in diverse sectors such as consumer products, construction, automobiles, and electronics. Two market dynamics have brought significant changes in the polyethylene industry: the increasing ability of mainland China to satisfy domestic demand and export PE goods, and the availability of abundant low-cost feedstock sources in countries such as the Middle East and North America. The many attributes of polyethylene have contributed to a significant increase of 42% in its capacity during the previous decade, leading to a notable expansion of 37 Million Metric Tons (MMT) [1],[2],[3].

Table 1. Polyethylene World Capacity, Production, Trade and Consumption in Million Metric Tons (MMT).

Year	Capacity	Production	Trade	Demand
2010	89.2	74.6	29.9	73.8
2011	92.2	76.8	32.1	76.6
2012	93.3	78	33.8	78.4
2013	97	81	35.7	81.7
2014	98.8	84	36.9	84.6
2015	101.5	88.2	39	88.3
2016	105.6	92.7	41.5	91.9
2017	111.4	96	44.2	96.4
2018	117	101.6	48.3	101.2
2019	119.9	104	50.9	104.3
2020	126.2	107.6	52.2	107.4
2021	133.9	113.2	51.4	113.2
2022	142.1	117.7	51.6	117.7
2023	148.8	122.3	51.9	122.3
2024	151.1	127.2	52.9	127.2
2025	154.2	132.1	54.8	132.1

China is expected to increase its polyethylene consumer base, adding 12-13 million metric tons of new capacity by 2025. However, mainland China will continue as a significant importer due to insufficient production capacity. Cracker outages in Europe have led to product shortages, and the competitiveness of export-oriented facilities in the Middle East is expected to cause restructuring. Global polyethylene consumption is projected to grow by 4.2% per annum. The High-Density Polyethylene (HDPE) market has grown by 4% in the past decade due to its cost advantages and diverse features. The COVID-19 pandemic has boosted HDPE usage in films and sheets, while demand grew by 2.9% in 2019-20. The market is influenced by feedstock availability and cost, with the Middle East being the most significant exporter. Pricing is influenced by supply and demand factors, with a cyclical pattern expected. Southeast Asia's HDPE capacity is 6% of global capacity, with 9% of global demand. The region is a net importer since 2015, with a projected 2.3% growth in demand from 2020 to 2025. Primary nations influencing supply and demand include Thailand, Malaysia, Singapore, Indonesia, the Philippines, and Vietnam [1].

Furthermore, the Film and Sheet application accounts for the largest share, comprising up to 48% of Southeast Asia's polyethylene consumption by end-use, followed by blow molding and injection molding, respectively as shown in Fig. 3 [1].

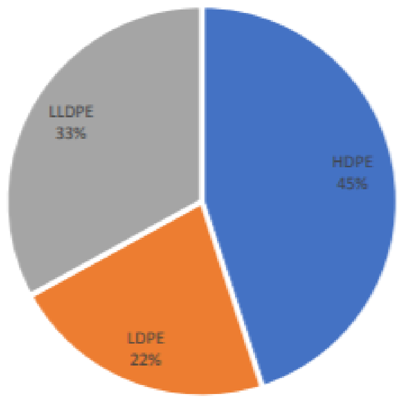

Fig. 1. World Consumption of Polyethylene by Product [1].

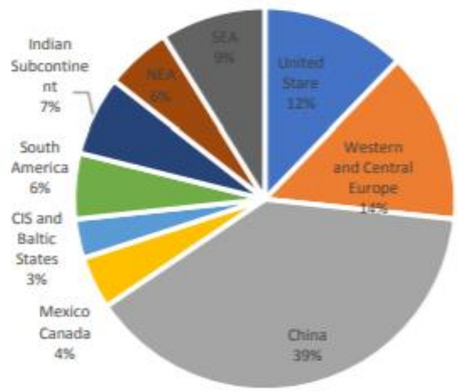

Fig. 2. World Consumption of Polyethylene by Region [1].

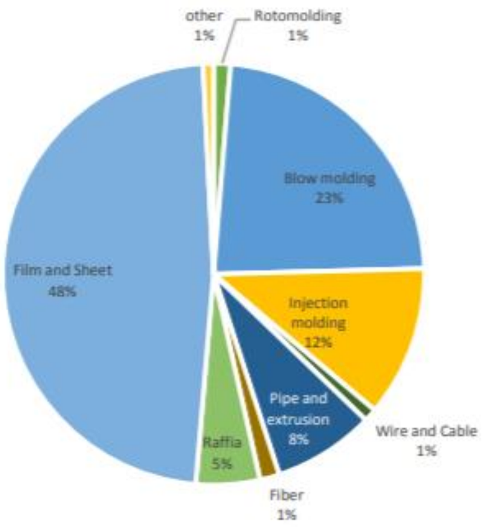

Fig. 3. Southeast Asia Consumption of Polyethylene by End-use [1].

DOI:10.4186/ej.2024.28.12.79

Pricing is a crucial aspect of business, impacting financial stability, profit margins, market position, competition, and customer perception. It is essential to understand the importance of pricing and its efficiency. Porter's 5 forces principle emphasizes the importance of pricing across multiple dimensions [4], while the 4Ps concept emphasizes the role of price in generating effective product strategies [5].

This paper focuses on the impact of price on Polyethylene (PE) products, examining marketing principles and the 5-force analysis. Factors such as production capacity, operational efficiency, plant shutdowns, maintenance activities, and unexpected disruptions can constrain supply, putting upward pressure on prices. Demand-side factors like end-user applications and industrial requirements also shape market dynamics. The intense competition among PE product rivals intensifies due to the presence of additional competitors. Supply-side factors are subject to significant volatility, directly impacting PE prices. The cost of switching is an important factor, as PE is a commodity with limited differentiation. The research uses a forecast model to focus on pricing and promotion strategies, highlighting the importance of effective promotional activities and improved predictive abilities for success.

HDPE, renowned for its high tensile strength and impact resistance, is ideal for applications requiring durability, such as containers, pipes, industrial products, and everyday commodities. In 2020, HDPE accounted for 45% of total consumption, a commodity that constitutes the largest share of consumption as illustrated in Fig. 1. Due to data availability and the location of the company case study in Southeast Asia. So, a predictive model will be developed to estimate the market price of High-Density Polyethylene (HDPE) film grade in the Southeast Asia (SEA) market.

2. Business Pain Point Analysis

Market analysis is crucial in the process industries, involving selling prices, product demands, market volumes, and processing networks. It involves life cycle curves, floor price and margin components, and price elasticity-demand correlations. Market evaluation is challenging due to factors like resources, technology advancements, production costs, global rivalry, and macroeconomic issues. Three case studies demonstrate the intricacies of market appraisal.

Chemical and plastics prices can fluctuate without notice, with crude oil costs impacting petrochemical product prices. Consumers seek predetermined price frameworks to enhance risk capital efficiency, maximize loan leverage, minimize earnings fluctuations, restrict capital investment, preserve cash flow, expand businesses, and take advantage of tax advantages. Based on data from ICIS, the spot market price is depicted in Fig. 3 and displays the world pricing history of HDPE from 2012 to 2022, while Fig. 4 displays only the Southeast Asia market.

Fig. 3. World Historical Market Price of HDPE.

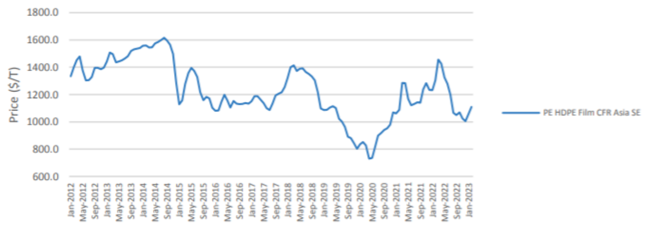

Fig. 4. Southeast Asia Historical Market Price of HDPE.

Several events have significantly impacted HDPE prices on a macro scale. Geopolitical factors, such as conflicts in the Middle East and the trade war between the US and China since 2018, have played a role. Natural disasters and accidents, like the US polar vortex in January-February 2019, and the obstacle at the Suez Canal in 2021, which directly affect HDPE trade flow have also influenced prices. Additionally, the commercial startup of PE plants in various regions, particularly in China, or plant shutdown events and regional arbitrage occurring multiple times a year, have affected HDPE prices.

In times of high volatility, businesses need precise price estimates and budget plans to adapt to economic changes. However, in uncertain situations, these require accurate forecasts and forecasts. To address this, organizations should increase the frequency of their projections and incorporate changes. Research shows that when market volatility increases, management's focus shifts to timely re-forecasts and dynamic projections. This provides valuable information for decision-making. For companies facing significant fluctuations and unpredictability, frequent analysis and re-forecasting are essential. An efficient budgeting/forecasting system, whether specialized software or Excel, is crucial for adaptability and capabilities. Enhanced predictive abilities give organizations a competitive edge by identifying challenges and prospects, aiding decision-making, and enabling swift responses.

Business analysts use linear regression models and industry experts to make monthly projections of HDPE commodity prices, with a three-month projection horizon. They gather data from monthly meetings to discuss market trends, demand, supply, end-user demand, and geopolitics. This comprehensive perspective helps establish a finalized forecast approximately six months in advance, providing a comprehensive view of the petrochemical supply chain.

DOI:10.4186/ej.2024.28.12.79

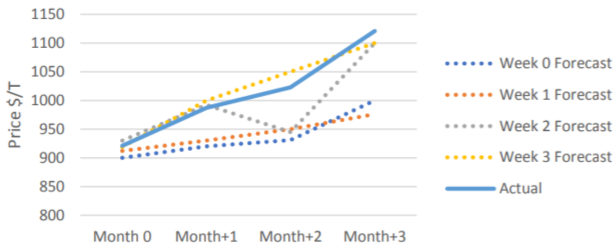

Fig. 5. Current business practice price forecast.

Pricing is a pivotal aspect of business, influencing revenue, profitability, market positioning, competitiveness, and customer perception. Effective pricing strategies enhance supply chain visibility, streamline workflows, increase profits, reduce costs, and mitigate risks, thereby improving overall business performance.

Price forecasting enables sales and marketing teams to develop strategic sales plans, effectively influencing market sentiment to maintain or boost sales volumes, ultimately aiming to increase business income and enhance the company's competitiveness. Additionally, production planning utilizes price forecasting to gauge market sentiment, optimizing inventory levels to timely support sales and marketing efforts.

This process not only enhances sales and marketing aspects but also improves financial and production facets. Therefore, pricing plans, sales volume plans, raw material plans, and production plans should be aligned. Currently, the raw material sourcing and production planning teams adjust their plans within a five-month scope or 20 weeks, while the sales volume plan is designed to capture market opportunities three months or 12 weeks in advance. To align these plans along with the polyethylene supply chain, it is crucial to have a price forecast that is neither too sudden, allowing the production and sourcing teams to optimize their plans, nor too prolonged, which could obscure market clarity for the sales and marketing teams.

As mentioned in Section 1, HDPE holds the largest share among polyethylene products, particularly in the film grade. HDPE film is considered a general grade and is commonly used as a reference price for other specialty products such as injection grade, pipe grade, or molding grade. Its status as a general grade makes data availability for this segment accessible and sufficient for complex models which require a big database for analysis.

This research aims to develop an HDPE product price forecast using deep learning techniques and operational process adjustments. It can identify suitable deep learning methodologies for predicting petrochemical pricing, provide recommendations for various sectors, and deliver use cases that provide a competitive edge. The research also aims to enhance understanding of determinants impacting petrochemical pricing in a volatile, uncertain, complex, and ambiguous (VUCA) global environment. A 16-week predictive model will estimate the market price of 'High-Density Polyethylene commodity film grade' in Southeast Asia, specifically in terms of 'Dollar per Metric ton'.

2.1. Market Analysis

The primary aim of this paper is not to analyze the business itself but to emphasize the impact and significance of price, examining various marketing principles. PE prices are influenced by production capacity, operational efficiency, plant shutdowns, maintenance activities, and unexpected disruptions, which can constrain supply and increase prices. Demand-side factors, such as end-user applications and industrial requirements, also shape market dynamics.

The 5-force analysis concludes that competition among PE product rivals is intense, with the industry experiencing general expansion. The presence of additional competitors intensifies rivalry as each competes for market share. As stated in section 1 on the Polyethylene market, upstream factors are highly volatile, directly impacting PE prices. Given PE's status as a commodity with limited differentiation, switching costs are significant; customers are more likely to switch vendors if these costs are high.

The forecast model developed in this thesis focuses on the pricing and promotion strategies of the 4Ps. Businesses aim for high profitability by selling products at optimal times and implementing effective promotional activities. Improved predictive abilities provide a competitive edge by identifying challenges, aiding decision-making, and ensuring timely forecasting.

3. Literature Review

3.1. Economic Theory

Economics is a social science that studies the creation, distribution, and utilization of commodities and services, as well as decision-making processes [6].

3.1.1. Supply, demand, and price

The law of supply and demand is explained, highlighting how price changes affect resource availability and desirability. An increase in price leads to a significant increase in supply, while a decrease results in a decrease in demand.

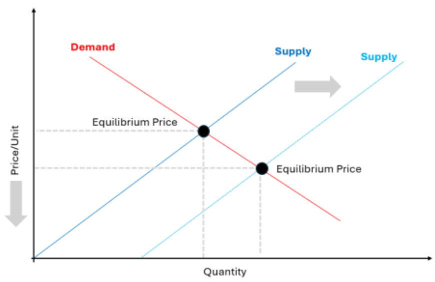

Fig. 5. Relationship between Demand, Supply and Price.

a) Price theory is a fundamental concept in microeconomics, suggesting that the optimal price for a commodity or service is determined by the interplay between supply and demand. Equilibrium refers to a state of balance or stability, where the price at which demand and supply are equal is the optimal price. If the price exceeds a threshold, customers may choose to abstain from purchasing, leading to an abundance of supply and potential producer price decreases. Conversely, if the price is excessively low, demand may exceed the available supply, leading to an increase in pricing. The clearing price, or ideal price, considers both supply and demand.

Price elasticity refers to the market's responsiveness to price changes, operating in both directions. It captures how consumers respond to price changes, while it quantifies how producers react to price fluctuations. Corporations prioritize price elasticity of demand when formulating sales strategies. Supply and demand do not always respond proportionally to price changes. High price elasticity of demand results in more pronounced variations in demand, while essential items have a relatively low price elasticity, as individuals cannot readily forgo them.

b) Price Trend is the direction of prices moving based on their historical trajectory, consisting of high and low points. It is determined by the movement of peaks and troughs, which can be upward, downward, or sideways.
    - Uptrend consists of progressively higher peaks and troughs. Ascending peaks and ascending troughs.
    - Downtrend consists of declining peaks and troughs. Declining peaks and declining troughs.
    - Sideways trend, also known as consolidation, refers to a situation when prices remain within a horizontal range without any significant upward or downward movement.

3.1.2. Economic indicators

Economic indicators offer a comprehensive overview of economic activity, influencing stocks, employment, and markets. They serve as reliable predictors of future economic conditions, shaping market movements and investment choices [7].

a) GDP is a crucial economic indicator, valuing the aggregate market value of final goods and services produced annually. Investors and analysts often focus on advanced and preliminary reports, as GDP is a trailing indicator.[8]
b) PMI was defined as a metric indicating economic trends in manufacturing and service sectors. It serves as a diffusion measure, providing data on prevailing and prospective business conditions to corporate decision-makers, analysts, and investors [9].
c) Production capacity is the maximum amount of goods a manufacturing facility can produce within a specific timeframe. It aids manufacturers in strategic and tactical business decisions, enhancing efficiency [10].

3.1.3. Data science

IBM categorizes data into structured and unstructured formats, each gathered and stored differently in different databases, highlighting the importance of understanding these disparities for professionals working with data [11].

- Structured data is systematically structured, easily searchable, and can be accessed using SQL or Structured Query Language. It includes numbers or characters.
- Unstructured data is any data without a predefined structure, including emails, photos, video files, audio files, social network posts, and PDFs.

Variables in machine learning can be categorized into two main types: numeric and categorical [12].

- Categorical variables, which are qualitative traits, cannot be measured or quantified. Nominal variables represent names, labels, or categories without inherent order. Ordinal variables have values determined by the relative order of categories.
- Numeric variables, quantitative attributes, are measurable and can be continuous or discrete. Continuous variables can take on unlimited real values within a specified range, like height and age. Distinct variables, on the other hand, can only take on finite real values within a specific interval. Examples of discrete variables include household counts or judge scores. These categories help in understanding the data and its use in machine learning.

Programming uses a range of standardized data types to store various types of data in memory [13].

- Int is used for signed integers of unlimited length, while
- long stores and operates on integers with a larger range.
- Float is used for floating precision numbers and can represent values up to 15 decimal places.
- Complex data types are used for complex numbers.
- Strings are contiguous sequences of characters, with Unicode support.
- list is a powerful data structure that can concurrently store multiple forms of data.
- Tuples represent unchangeable data,
- Dictionaries are unordered collections of key-value pairs.

3.1.4. ETL process

ETL is a sequential computing process that involves extracting data, transforming it, cleaning it, and finally loading it into an output container, allowing data from multiple sources and destinations [14],[15].

a) Extract: Data extraction involves duplicating or transferring unprocessed data from its original sources to a designated area for temporary storage. Data management professionals can extract data from various sources, including structured and unstructured formats.
b) Transform: During the staging phase, unprocessed data undergoes data processing, including filtering, cleansing, de-duplication, validation, and authentication. Audits verify data accuracy and adherence to standards, while data subject to business or governmental regulations is secured. Data is reorganized into tables and stored in a designated data warehouse.
c) Load: The final stage involves loading the converted data into the warehouse, typically during non-peak hours with minimal activity on both source systems and the data warehouse.

3.1.5. Features engineering

Feature engineering converts raw data into features for prediction models, improving accuracy. Feature selection involves selecting relevant features to enhance model performance [16],[17].

a) **Feature Generation**: Consolidating data into a reliable source allows for transformations to refine representations, enhancing model performance. This process uses signal processing and technical indicator approaches to derive additional characteristics from the original data inputs, allowing for the application of various transformations to features.

- Signal processing is a crucial field in data science that involves extracting, analyzing, and manipulating signals and time-series data. Signals can be analogue or digital, and time-series data involves recording measurements at consecutive intervals [18]. Symbolic Aggregate Approximation (SAX) is a technique used to examine fluctuations in target variables and accurately represent current trends by dividing the target signal into intervals [19].
- Technical indicators, such as the Exponential Moving Average, Moving Average Convergence Divergence, and Bollinger Band, are heuristics used by traders to forecast future price changes based on historical data [20].

  The exponential moving average (EMA) is a statistical tool that generates purchase and sale signals by analyzing historical average divergences and crosses over a specific timeframe [20].

$$EMA = \left(P_{current} \times \frac{2}{N+1}\right) + EMA_{Previous} \times (1 - \frac{2}{N+1})$$

where P_current is the current price, and N is a number of time periods.

The MACD or Moving Average Convergence Divergence indicator, a lagging momentum indicator, demonstrates the relationship between two EMAs and can serve as a leading indicator for forecasting market changes [20].

$$MACD = EMA_{fast} - EMA_{slow}$$

where EMA_{fast} is price fast-moving period EMA, and EMA_{slow} price slow-moving period EMA.

Bollinger Bands are a technical technique developed by John Bollinger to indicate stock oversold or overbought, assessing upper and lower price boundaries in a market. They use the standard deviation of the target variable [21].

$$Bol_{Up} = MA(Tp, n) + m \times \sigma[TP, n]$$

$$Bol_{low} = MA(Tp, n) + m \times \sigma[TP, n]$$

$$TP = \frac{High + Low + Close}{3}$$

where Bol_{Up} is Upper Bollinger Band, Bol_{low} is Lower Bollinger Band., MA is Moving average, TP is typical price, n is a number in the smoothing period, m is the number of standard deviations, and $\sigma[TP, n]$ is the standard deviation over the last n period of TP

b) **Feature Selection**: Feature selection is the process of selecting a specific set of relevant features, such as variables or predictors, for the model [22].

Correlation testing is a crucial statistical method used by researchers to analyze data relationships, ensuring model fairness and preventing bias.

- Pearson's correlation coefficient measures the strength of a linear relationship between variables, with scores closer to the extremes indicating less variation and better linear connection, and closer to zero indicating insufficient data capture [23].

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where r is Pearson Correlation Coefficient, x_i is x variable sample, y_i is y variable sample, $\bar{x}$ is mean of values in x variable, and $\bar{y}$ is the mean of values in the y variable.

- Stationarity testing ensures constant mean and variance in time series, crucial for accurate test statistics and model selection, despite not being essential for estimating parameters in econometric models [24].

  The augmented Dickey-Fuller (ADF) test is a widely used unit root test used to determine the stationary of a time series with p-values smaller than 0.01 [24].

$$y_t = \alpha + \beta t + \emptyset y_{t-1} + e_t$$

Or

$$y_t = \alpha + \beta t + \emptyset y_{t-1} + e_t$$

  where y_t is data. It is written this way so we can do a linear regression of Δy_t against t and y_{t-1} and test if γ is different from 0. If γ = 0, then we have a random walk process. If not and -1<1+γ<1, then we have a stationary process.

  The Augmented Dickey-Fuller test allows for higher-order autoregressive processes by including Δy_{t-p} in the model. But our test is still if γ = 0.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots$$

  The null hypothesis for both tests is that the data are non-stationary.

- The cointegration test is a method used to determine the level of responsiveness of two variables to the same mean price over a specific timeframe. It is used to reject the null hypothesis and suggest cointegration of characteristics, with individuals exceeding the threshold being eliminated [25],[26],[27].

c) **Feature importance**: Feature importance is a method that assigns numerical values to each input feature in a model, with higher scores indicating greater influence on the predictive model for a given variable [28].

- The XGBoost algorithm's feature importance method, accessible via the xgbModel.feature_importances_ attribute, determines the relative significance of features, with features with a significance greater than 0.1 retained, resulting in a gain in accuracy [29].
- The Random Forest feature importance algorithm was used to examine potential features, retaining only those with an importance factor greater than 0.1, and discarding others [29].

3.1.6. Deep learning model

Deep learning is an AI technique that mimics the human brain's function by using numerous layers within a network. Deep neural networks (DNNs) are hierarchical structures of interconnected neurons, organized into multiple layers [30]. Unlike a basic ANN with a single hidden layer, DNNs have numerous hidden layers. Each layer receives activation from the previous layer and is connected to the next to form a network structure 31].

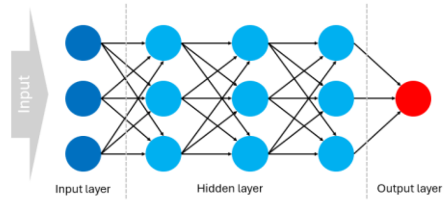

Fig. 6. General Neuron Network Model (NN).

a) Convolutional neural networks or CNN. CNNs use one-dimensional convolutional filters to capture local patterns or features at various points in an input sequence. Filters, or kernels, calculate weighted sums of surrounding elements [32].

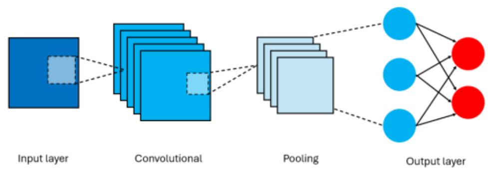

Fig. 7. Convolutional Neural Networks Model (CNN).

b) Recurrent neural networks (RNNs) are artificial neural networks designed for handling sequential data. They use a recurrent link to transmit information from previous stages to the present step, effectively capturing temporal dependencies between data points. The general structure of RNN. The LSTM cell is a widely recognized variation of RNN cells, integrating a memory mechanism. It can retain information for extended periods, discard irrelevant information, and modify content based on fresh input. The LSTM architecture consists of input, forget, and output gates, controlling information movement for efficient data capture [33].

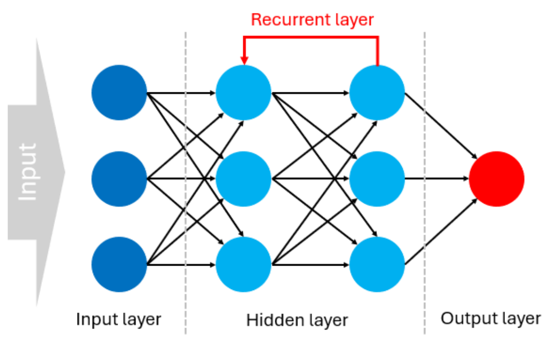

Fig. 8. Recurrent neural networks (RNNs).

DOI:10.4186/ej.2024.28.12.79

Table 2.

Authors, Year	Model											Input				Error						Business						
	Linear Regression	GARCH	ARIMA	ARFIMA	RNN	CNN	LSTM	encoder-decoder	Other Hybrid	Text Mining	Other	Market data	Economics data	Text data	Other	MAE	MAPE	MSE	RMSE	Coefficient	Other	Economic/Finance	Stock/Exchange	Oil/Gas	Gold	Petrochemical	Electricity	Other
Amral et al., 2007 [42]	x											x	x				x					x					x	
Atsalakis&Valavanis ,2009 [43]	x		x								x	x	x															
Baldi, 2012 [44]								x				x					x	x										
Bildirici&Ersin, 2014 [45]	x										x		x			x						x						
Boubaker et al., 2022 [46]				x		x	x					x	x				x		x					x				
H.Bukhari et al.,2020 [47]			x	x			x					x	x			x	x		x			x						
F. Chen et al., 2016 [48]						x						x					x		x			x						
Cho et al., 2014 [49]					x			x				x	x							x								x
Chung et al, 2014 [50]					x		x					x	x								x		x					x
Doering et al., 2017 [51]						x						x	x			x							x					
Elliott&Timmermann,2016 [52]	x												x			x	x					x						
Fischer&Krauss, 2018 [53]					x		x					x	x				x				x	x	x					
García-Martos et al., 2013 [54]		x	x									x					x							x				x
Hansen, 2016 [55]	x											x	x								x	x						
Hrasko et al., 2015 [56]	x		x								x	x							x				x					
Jiang, 2021 [57]					x	x						x	x	x		x	x		x				x					
Kaastra, 1996 [58]	x											x	x				x					x						
Kanwal et al., 2022 [59]									x		x	x	x			x							x					
Kohzadi et al., 1996 [60]			x								x	x				x			x			x						x
Kristjanpoller&Minutolo, 2015 [61]		x									x	x				x	x	x	x			x						
T.H. Le&T.N. Nguyen, 2023 [62]			x						x				x				x					x						
Lendasse et al., 2000 [63]											x	x						x				x						
Li, 2015 [64]					x							x				x	x					x						
X. Li et al., 2019 [65]					x	x				x				x	x			x							x			
Z. Li et al., 2022 [66]						x																						
Liu et al., 2017 [67]						x	x					x											x					
Q. Lu et al., 2010 [68]											x	x				x	x							x				
Peng et al.,2020 [69]											x	x					x							x				
Shi et al., 2022 [70]		x			x	x	x				x	x				x	x	x				x						
X, Shi et al., 2015 [71]						x	x	x							x			x										x
Md. Sunny et al, 2020 [72]	x		x	x			x				x	x						x					x					
Ma. Tapia et al, 2008 [73]	x										x	x	x			x	x					x						
Amit et al., 2017 [74]											x				x	x												x
Vuong et al., 2022 [75]			x				x					x	x			x		x	x				x					
Wang&Chiang, 2011 [76]										x				x			x											x
Wei&Zhen-gang, 2009 [77]			x								x				x			x									x	
Weldon et al.,2022 [78]					x	x	x	x	x		x																x	
Xiao et al., 2016 [79]									x		x	x			x	x	x	x									x	
Xiao et al., 2016 [80]									x		x	x			x	x	x	x									x	
Yu&Xu, 2014 [81]											x				x	x		x	x								x	
Zhao et al., 2018 [82]			x														x	x										
Zhao et al., 2020 [91]					x	x	x					x	x				x	x					x					
This research Proposed	**x**				**x**	**x**	**x**	**x**	**x**	**x**		**x**	**x**	**x**		**x**	**x**				**x**					**x**		

c) CRNNs are a hybrid neural network that combines the strengths of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks to capture spatiotemporal patterns in data, such as time series forecasting. The CNN extracts spatial features from the input sequence, producing a sequence of feature vectors, which are then passed into a recurrent layer to capture temporal dependencies and long-term trends [34].

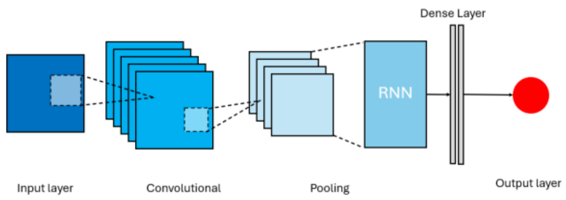

Fig. 9. Convolutional-recurrent neural networks (CRNNs).

d) Autoencoders are neural network models that use an encoder-decoder architecture to acquire effective data representations. They are used in time series forecasting to collect compressed past data and make future forecasts. There are multiple options for constructing an encoder and decoder, with three types suggested: CNN, LSTM, and ConvLSTM networks [35],[36].

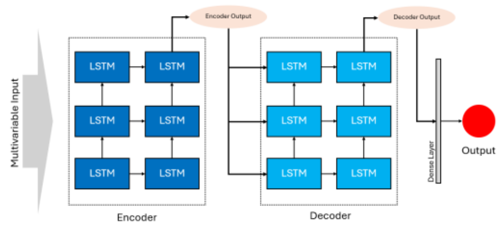

Fig. 10. Autoencoder.

3.1.7. NLP model

NLP Model or Nature language processing is a subfiel of artificial intelligence that facilitates the comprehension, interpretation, and manipulation of human language by computer systems.

a) FastText: The skipgram model, a fast method for representing words as character n-grams, is used for quick training on large corpora. The authors evaluated their word representations on nine languages and found state-of-the-art performance [37].
b) GloVe: GloVe is an unsupervised learning algorithm that uses global word-word co-occurrence statistics to create vector representations of words, revealing interesting linear substructures in the word vector space [38].
c) ConcepNetNumberBatch: ConceptNet is a knowledge graph connecting natural language words and phrases with labeled edges from expert-created resources, crowdsourcing, and games, improving natural language applications by providing deeper understanding than distributional semantics alone [39].
d) Sent2Vec: Sent2Vec is an efficient, unsupervised method using C-BOW-inspired techniques for sentence embeddings, outperforming competitors like SkipThought in supervised evaluations but struggles on similarity tasks [40].
e) Universal Sentence Encoder: The study presents efficient and accurate sentence embedding models for transfer learning in Natural Language Processing tasks, outperforming word-level transfer and demonstrating good performance with minimal supervised training data [41].

3.1.8. Gaps in the current literature

This section refers to the literature review summarized in Table 2, which provides an overview of research related to price forecasting using deep learning models. The review is divided into four parts: the predictive model employed, the type of data input, methods for monitoring model performance, and the focused business area.

Predictive Models: Various techniques are used to predict time series values, each with its limitations. Fundamental analysis, which relies on a forecaster's expertise, has declined in popularity due to its inefficiency. Statistical techniques like generalized autoregressive conditional heteroskedasticity (GARCH) and autoregressive integrated moving average (ARIMA) use predefined mathematical frameworks and statistical assumptions to analyze time series data. However, these methods may fail to capture the nonlinear patterns and complex interactions present in real-world time series. Artificial neural networks (ANNs) can automatically identify hidden patterns in historical data using machine learning techniques, but they depend on feature engineering and face challenges with model complexity when handling large datasets. Consequently, there is a need for more advanced price prediction techniques.

LSTM networks, a type of recurrent neural network (RNN), are highly effective in capturing relationships in time series data and have become the preferred method for price prediction, outperforming traditional statistical and machine learning approaches. Convolutional-recurrent neural networks (CRNNs), which combine the strengths of convolutional neural networks (CNNs) and RNNs, are also commonly used for price predictions due to their ability to analyze spatio-temporal patterns in time series data.

Data Input Types: Numerical data, such as economic indicators, prices, and costs, are easily manageable and provide detailed information that is straightforward to interpret and use in calculations. Numerical data is readily available from various free sources and can be easily stored using common tools like Excel. As a result, most published research utilizes numerical data for price prediction models. Numerical data includes economic indicators and general market data. Only a small number of studies use text mining models for price forecasting due to the time and cost involved, as well as concerns about model accuracy. The complexity and duration of model runs increase with the quantity of report input, and subscriptions to news and insight reports can be expensive. Given the slight difference in advantages between text-based and numerical-based models, most researchers prefer numerical data.

Model Performance Evaluation: Various error measurements are used to evaluate model performance, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). MSE calculates the mean squared deviation between predicted and observed data, focusing on significant errors, and is chosen when large errors are more critical. RMSE provides a more straightforward explanation of errors and is selected when the error scale needs to match the target scale. MAE measures the mean absolute difference between predicted and observed values and is more robust to outliers. MAPE represents the percentage difference between estimated and actual values, making it easy to interpret as a percentage and suitable for forecasting and analyzing errors based on percentages. SMAPE (Symmetric Mean Absolute Percentage Error) is also used for similar purposes.

Business Area: Most literature focuses on finance and the stock market, which can be segmented into various business sectors. However, these works often do not specify particular sectors. Several studies aim to predict macroeconomic indicators like GDP, while others focus on the prices of gold and crude oil. The complexity of forecasting gold and crude oil prices is heightened by their significant susceptibility to geopolitical factors.

This review highlights both the uniformity and particularity in studies on deep learning models for price forecasting. Uniformity is observed in similar target variables and prediction processes, while distinctiveness arises from the diverse models with unique architectures. Although the field has made rapid progress, much remains to be explored.

4. Research Methodology

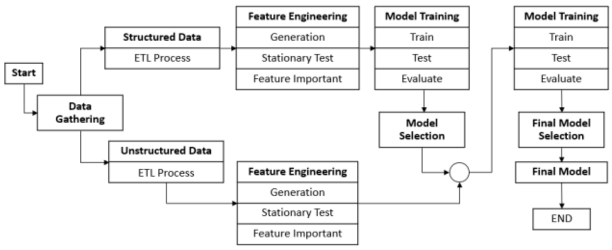

Fig. 11. Overview Process.

This study focuses on the use of deep learning models to analyze the price of HDPE in the Southeast Asian market. The data is analyzed using economic theory, which suggests that the optimal price for a commodity or service is determined by the market's equilibrium between supply and demand. The research aims to analyze the price of HDPE in the region by collecting data relevant to chemical prices and economic indicators.

Data preprocessing is crucial in data analysis, transforming raw data into understandable formats and ensuring data quality before applying machine learning or data mining algorithms. Feature engineering transforms raw data into features that represent the problem better for predictive models, improving model accuracy. Feature generation involves signal processing and technical indicator approaches to extract additional characteristics from data inputs.

Model development and architecture selection are divided into four phases: data exploration and gathering, data processing, model development and model architecture selection, and model implementation. Machine learning training involves acquiring dataset statistical characteristics, applying them to new situations, and evaluating models through testing. Performance criteria such as MAE, MAPE, and SMAPE provide an impartial perspective.

A comparison of five text deep learning models, FastText, GloVe, NumberBatch, Sent2Vec, and Universal Sentence Encoder, will be conducted to determine the most suitable model for a business. The evaluation will be based on resources and time consumption, with a single Deep Learning model chosen for text encoding and incorporated into the selected deep learning model for structured data.

In the final phase, the research examines the existing business process and identifies multiple problems. It proposes a method to enhance the business process by improving forecasting time and accuracy. The proposed deep learning model is recommended to boost corporate competitiveness in volatile markets and increase business revenue by capitalizing on captured opportunities.

4.1. Data Preparation

In this paper, we refined the scope of our inputs by applying economic theory. As discussed, price theory—a fundamental concept in microeconomics—suggests that the optimal price for a commodity or service is determined by the market equilibrium between supply and demand. The objective of this thesis is to analyze the price of HDEP in the Southeast Asian (SEA) market. Consequently, our focus is on collecting data directly relevant to chemical prices and economic indicators, such as crude oil prices, GDP, and PMI across various regions. Based on data availability and the applied economic theory, the primary dataset for this model will encompass a five-year period from 2014 to 2019. This dataset includes various polyethylene (PE) products such as HDPE, LDPE, and LLDPE across different applications, as well as crude oil prices, country GDP, and PMI.

Missing value imputation (MVI) is a method used to address incomplete dataset issues, particularly when data samples have missing attribute values [83]. Deep Learning models can remove missing data, fill with statistical values, and ensure data continuity by replacing missing data points with the latest known value until a new observation is available for further analysis or modeling. Forward fill involves using regression or K-Nearest Neighbors models, but these methods may sometimes lead to overfitting of the data.

Table 3. Data Category for Price Forecast Model.

Type	Category
Structured	Commodity Price
	Industry Operating Rate
	Industry demand
Unstructured	Petrochemical News

4.2. Feature Engineering Process

The Feature Engineering Process is divided into three phases: Feature Generation, Feature Selection, and Feature Importantity. Feature Generation involves extracting, interpreting, and manipulating signals and time-series data, which can be either analogue or digital. The SAX method is used to analyze fluctuations in target variables, while indicators like the Exponential Moving Average, Moving Average Convergence Divergence, and Bollinger Band are used to forecast future price changes. Feature Selection involves Pearson's correlation coefficient, unit root tests, and cointegration function to determine the strength of linear relationships between variables. Feature importance is determined using XGBoost and Random Forest algorithms, with higher scores indicating greater influence on the predictive model. Both methods aim to improve the accuracy of a model. The process is divided into three phases, with the most important feature identified at the end.

4.3. Model Input

The model architectures are consolidated under a single interface, ensuring equal evaluation. Input data is a panda DataFrame with "date" and "target" columns, while output data is a pandas Series with a date index. Deep-learning models are sequence-to-sequence, taking a fixed sequence length as input and predicting a fixed output length.

4.4. Training Regime

Machine learning training involves acquiring the statistical characteristics of a dataset and applying them to new situations. Evaluating models through testing is crucial to determine their robustness. Machine learning models are trained using a suitable algorithm and training data, which are divided into training and testing data. Data is often divided randomly, but time series data cannot be partitioned arbitrarily. The optimal division for training, validation, and test sets is 70%, 15%, and 15% [84],[85].

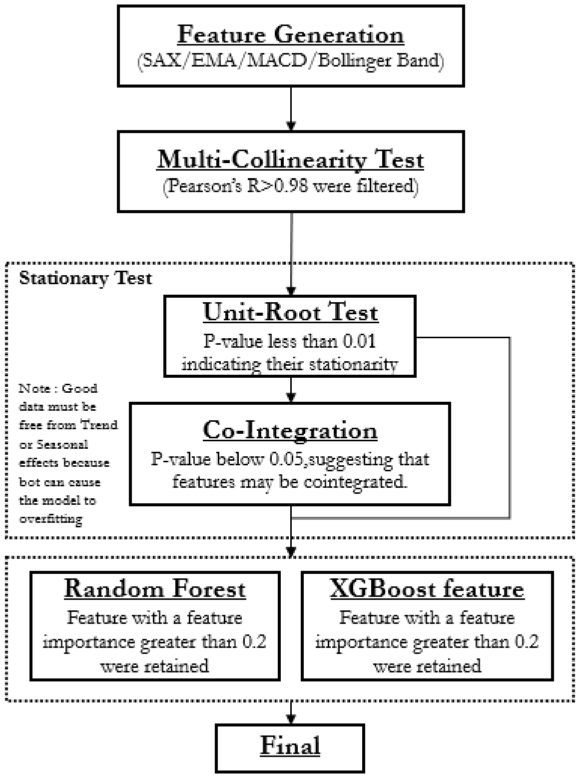

Fig. 12. Feature Selection Framework.

Table 4. Train-Test-Validate periods.

Ratio	Time Periods
100%	1 January - 31 December 2022
70%	1 January - 12 September 2022
15%	13 September - 6 November 2022
15%	7 November - 31 December 2022

4.5. Model Optimizer

Deep learning optimizers, such as Stochastic Gradient Descent, Adam, and RMSprop, work to minimize model error and improve performance by fine-tuning neural network parameters. These optimizers employ distinct rules, learning rates, and momentum strategies to find optimal model parameters. The Adam optimizer, which updates neural network weights during training, integrates the advantages of AdaGrad and RMSProp optimizers, enabling dynamic adjustment of learning rates and faster convergence [86],[87],[88].

4.6. Performance Evaluation

Training methodologies evaluate performance by making predictions on a test dataset and comparing them to known values. However, other methods also exist. To accurately analyze performance, a criterion must provide an impartial perspective, avoiding influence from specific evaluation circumstances [89],[90].

$$MAE = \frac{\sum_{i=1}^{n}|F_t - A_t|}{n}$$

where A_t is the actual value and F_t is the forecast value. The value of this calculation is summed for every fitted point t and divided again by the number of fitted points n.

The Mean Absolute Percentage Error (MAPE) is a widely used metric for evaluating time series data performance due to its dimensionless nature, scale-independent nature, simplified interpretation, and absolute values that prevent positive and negative errors from nullifying each other, making it a valuable tool for data analysis.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

where A_t is the actual value and F_t is the forecast value. Their difference is divided by the actual value A_t. The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points n.

$$SMAPE = \frac{100}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

where A_t is the actual value and F_t is the forecast value.

The absolute difference between A_t and F_t is divided by half the sum of absolute values of the actual value A_t and the forecast value F_t. The value of this calculation is summed for every fitted point t and divided again by the number of fitted points n.

4.7. Deep Learning Architecture Selection

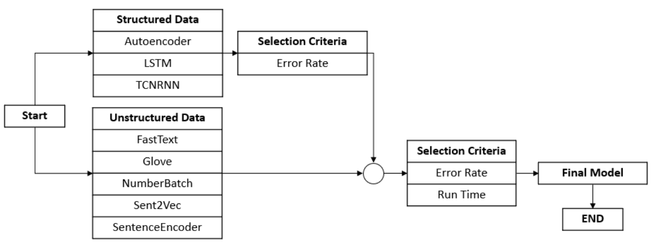

Fig. 13. Deep Learning Architecture Selection Framework.

The selection framework for text models involves comparing five deep learning models: FastText, GloVe, NumberBatch, Sent2Vec, and Universal Sentence Encoder. The evaluation is based on space and time consumption, with a single deep-learning model chosen for text encoding to capture market sentiment. Two scenarios were tested for text encoders, with the hypothesis that models would improve average runtime. The structured data model will be divided into three models: EncoderDecoder, LSTM, and TCNRNN, which are considered advanced for price prediction. The model's accuracy is primarily evaluated based on business requirements.

LSTM (Long Short-Term Memory) networks, Autoencoders, and TCNRNN (Temporal Convolutional Neural Recurrent Neural Networks) were selected for the price forecasting model due to their distinct advantages in handling time series data and capturing complex patterns. LSTM models are extensively utilized for time series forecasting, as highlighted by Shi, Jain, and Narasimhan [70]. Their proficiency in capturing long-term dependencies renders them particularly effective for this application. For instance, Bukhari, Fischer, and Krauss have employed LSTM models for financial market forecasting [47], [53]. Similarly, Sunny et al. and Vuong et al. have also leveraged LSTM models for stock price forecasting [72], [75]. Additionally, the Encoder-Decoder architecture offers various configurations for constructing the encoder and decoder components as Roy et al. [35] used it for stock prediction, and Li & Becker used it for electricity price prediction [36]. Lastly, TCNRNN, which combines temporal convolutional networks and recurrent neural networks, has shown promise in capturing both short-term and long-term dependencies in time series data. This hybrid approach can be beneficial for stock index forecasting by Zhao et al. [91].

This research will use Microsoft Excel as the primary database for data storage, table and graph creation, and summary generation. Excel is a popular choice for data analysis due to its simplicity, familiar layout, and wide range of capabilities. It can process large volumes of data, execute computations, and generate charts, graphs, and tables. Visual Studio is a comprehensive solution that enables the entire development cycle within a single environment. Python is a versatile programming language suitable for web and software development, machine learning, and AI applications. Its straightforward syntax and numerous accessible libraries make it accessible to users. NumPy is a library of coding that simplifies the manipulation of large arrays and matrices, making it useful for deep learning calculations. SciPy is designed for scientific computations on large datasets, with modules for array optimization and linear algebra operations. Pandas offers tools for data manipulation and analysis, with robust data structures for numerical tables and time series analysis. TensorFlow is a library for differentiable programming, offering tools for deep learning, machine learning, and neural networks. Lastly, NLTK is a collection of text-processing modules and Python platforms for human language data processing and analysis.

5. Results and Analysis

The feature engineering results from correlation, unit root, and co-integration tests are presented, with feature names arranged in a specific pattern to align with data arrangement since the ETL process. The example "pbh_sea_hdpe_price_unspecified" can be interpreted as the price of HDPE of SEA from the source pbh.

5.1. Unit Root Test Result

A unit root is a metric used to measure stationarity in a time series model. If it doesn't exist, the null hypothesis is rejected, indicating non-stationarity. If it does, the p-value is small, indicating stationarity [24].

Table 5. Unit Root Test Results.

	Feature name	p value	adf Stat
1	mof_japan_pvc_import_volume_unspecified	0	-7.944
2	meti_japan_meg_output_vol_unspecified	0.00001	-5.203
3	venezuela_refinery_processing_gain_mbd	0.00001	-5.257
4	gabon_refinery_processing_gain_mbd	0.00001	-5.146
5	mof_japan_pp_copol_export_volume_unspecified	0.00002	-4.976
6	meti_japan_pp_output_volume_unspecified	0.00002	-5.047
7	meti_japan_butadiene_stcks_vol_unspecified	0.00005	-4.837
8	brunei_refinery_processing_gain_mbd	0.00006	-4.786
9	singapore_refinery_processing_gain_mbd	0.00009	-4.696
10	lldpe_isc_inventory_change	0.00011	-4.641
11	ldpe_mde_operating_rate	0.00018	-4.522
12	congobrazzaville_refinery_processing_gain_mbd	0.00021	-4.4
13	meti_japan_meg_stcks_vol_unspecified	0.0003	-4.394
14	meti_japan_propylene_output_volume_unspecified	0.00035	-4.358
15	meti_japan_ethylene_output_vol_unspecified	0.00063	-4.211
16	ldpe_isc_operating_rate	0.00076	-4.161
17	meti_japan_pe_output_volume_unspecified	0.00086	-4.132
18	ldpe_mde_imports	0.00089	-4.122
19	nigeria_ngpl_mbd	0.00129	-4.024
20	meti_japan_butadiene_output_vol_unspecified	0.00141	-4.000

5.2. Co-integration Test results

A cointegration analysis is used to determine the relationship between the target variable and other features. A null hypothesis suggests no cointegration, while a low p-value suggests cointegration. A stationary time series is rejected if the p-value is small [25].

Table 6. Co-integration Test results.

	Feature name	p value	adf Stat
1	china_pe_hdpe_film_price	0	-6.8438
2	asia_se_ethylene_price	0.00001	-5.6166
3	asia_ne_ethylene_price	0.00016	-5.0089
4	mof_japan_pvc_export_volume_unspecified	0.00192	-4.3797
5	china_qend_fxrate_usd	0.00648	-4.0313
6	natural_gas_spot_price	0.00761	-3.982
7	asia_se_styrene_price	0.00877	-3.9385
8	china_pe_lldpe_film_price	0.0101	-3.8948
9	global_ethylene_index_unspecified	0.01118	-3.8617
10	weekly_natural_gas_futures_contract_1_dollar_million_btu	0.01557	-3.7541
11	asia_ne_styrene_price	0.02326	-3.6185
12	china_styrene_monomer_price	0.02825	-3.5498
13	china_meg_price	0.03375	-3.4847
14	glob_economic_policy_uncertainty	0.03846	-3.4368
15	weekly_natural_gas_futures_contract_4_dollar_million_btu	0.04316	-3.3928
16	weekly_natural_gas_futures_contract_2_dollar_million_btu	0.04362	-3.3884
17	kuwait_crude_oil_including_lease_condensate_mbd	0.04481	-3.3781
18	weekly_natural_gas_futures_contract_3_dollar_million_btu	0.04974	-3.3376
19	mof_japan_lldpe_import_volume_unspecified	0.05665	-3.2867
20	kuwait_crude_oil_ngpl_and_other_liquids_mbd	0.05778	-3.2789

5.3. Correlation Test Results

The Top 15 Positive Correlation Test Results and Top 15 Negative Correlation Test Results are presented in Table 7 and Table 8 respectively.

Table 7. Top 15 Positive Correlation Test Results.

	Positive Correlation	Correlation
1	china_pe_hdpe_film_price	0.99093
2	asia_se_pe_lldpe_film_price	0.9648
3	china_pe_lldpe_film_price	0.95745
4	asia_se_ethylene_price	0.92898
5	asia_ne_styrene_price	0.91927
6	asia_se_styrene_price	0.91819
7	global_ethylene_index_unspecified	0.9171
8	china_styrene_monomer_price	0.91465
9	asia_se_ps_gpps_price	0.90909
10	china_ps_gpps_price	0.90536
11	china_pe_lldpe_c6_metallocene_price	0.90445
12	global_pp_index_unspecified	0.90146
13	asia_ne_ethylene_price	0.89535
14	global_ldpe_index_unspecified	0.89445
15	asia_se_pp_flat_yarn_raffia_price	0.89237

Table 8. Top 15 Negative Correlation Test Results.

	Negative Correlation	Correlation
1	kuwait_crude_oil_including_lease_condensate_mbd	-0.84163
2	united_arab_emirates_crude_oil_including_lease_condensate_mbd	-0.83215
3	korea_republic_of_qend_gdp_usd	-0.82508
4	hdpe_nea_production	-0.8204
5	kuwait_crude_oil_ngpl_and_other_liquids_mbd	-0.81335
6	kuwait_total_petroleum_and_other_liquids_mbd	-0.81077
7	united_arab_emirates_total_petroleum_and_other_liquids_mbd	-0.81068
8	united_arab_emirates_crude_oil_ngpl_and_other_liquids_mbd	-0.80984
9	vietnam_refinery_processing_gain_mbd	-0.79735
10	hdpe_nea_total_supply	-0.78885
11	iraq_ngpl_mbd	-0.78464
12	lldpe_isc_domestic_demand	-0.7806
13	glob_economic_policy_uncertainty	-0.77952
14	australia_qend_gdp_usd	-0.77762
15	china_qend_gdp_usd	-0.77753

5.4. Feature Important Test Results

Table 9 identifies 9 important features, classified into three levels: Level 1 (highly important), Level 2 (modestly important), and Level 3 (important).

Table 9. Feature Important Test Results.

	Features	Level
1	cushing_ok_crude_oil_future_contract_2_weekly	Level 1
2	natural_gas_futures_contract_3_weekly	Level 1
3	natural_gas_futures_contract_4_weekly	Level 1
4	cushing_ok_crude_oil_future_contract_1_weekly	Level 2
5	cushing_ok_crude_oil_future_contract_4_weekly	Level 2
6	pp_biaxially_oriented_film_cfr_fe_asia_weekly_low_price	Level 2
4	hdpe_film_cfr_s_asia_mavg_high_price	Level 3
5	hdpe_film_fob_middle_east_netback_wavg_close_value	Level 3
9	pe_hdpe_bm_cfr_chi_all_ogs_spot_0_8_wks_fmrw_mid	Level 3

5.5. Models Forecasting Result

This section presents HDPE price forecast results using various methods, including conventional human forecasting, structured data-based forecasting, and a mix of structured and unstructured data. The model generates 12 future forecast values and requires 52 iterations or weeks for a single model. The study tested three deep learning architectures for structured data and selected the most optimal architecture for use with five natural language processing models.

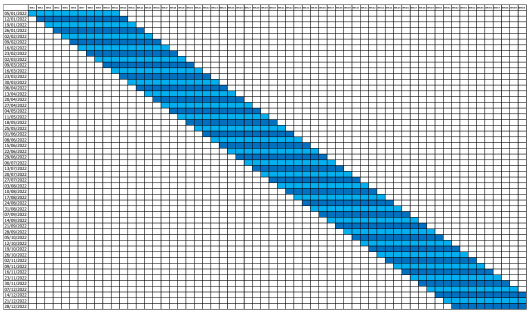

Fig. 14. Illustration of Forecast Results.

a) Baseline Model

The forecast process for HDPE is based on three main factors: market price, market insights from reliable sources, and internal discussions across the supply chain. The forecast results are updated monthly and based on market conditions.

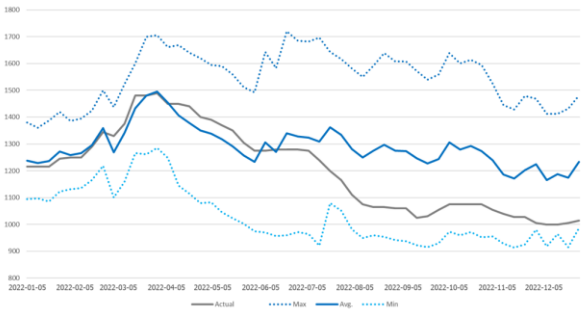

Fig. 15. Baseline Model Forecast Result.

The structured data model will consist of three advanced deep learning models: EncoderDecoder, LSTM, and TCNRNN. The model will be evaluated based on accuracy using parameters MAE, MAPE and SMAPE. Quantitative and qualitative evaluations will be conducted to identify the most effective combination of model and data inputs. Cross-validation will be performed for each combination. Performance evaluation will involve comparing the model to baseline models to verify its legitimacy and ensure accurate prediction of outcomes.

b) LSTM

The LSTM architecture, a variation of RNN cells, is used in this research to forecast polymer product prices in Southeast Asia. The parameters in Table 10 were developed from literature and Adams optimizer, based on time series data.

Table 10. LSTM Model Parameter for Polymer Price forecast in Southeast Asia.

Parameter Name	Value
Layer	2
Hidden	32
Dropout	0
EPOCHS	50
Batch Size	64
Learning Rate	0.025
Weight Decay	1.25
Batch Size	128

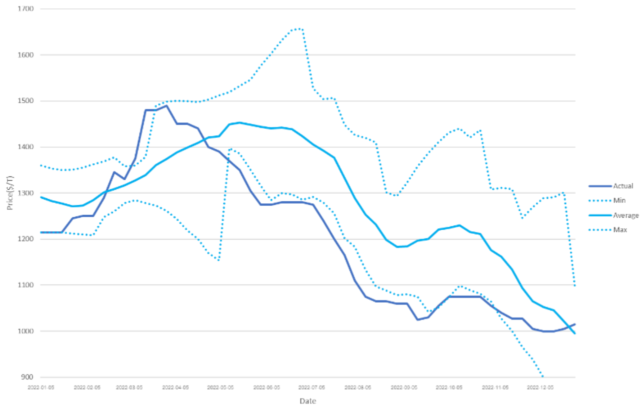

Fig. 16. Result of Polymer Price Forecast in Southeast Asia by LSTM.

The LSTM deep learning model predicts HDPE product prices from January 2022 to December 2022, with dark blue dotted lines representing the maximum, light blue dotted lines representing the minimum, and solid blue lines representing the average. The average forecast results have a low correlation with the actual market price, with MAE, MAPE and SMAPE values of 75, 20.1%, and 19.9% respectively.

c) Autoencoder

Autoencoders are neural network models using an encoder-decoder architecture for time series forecasting. They convert historical data into a feature vector, which the decoder uses to produce an output. This model is used for Southeast Asia polymer product price forecasting. The parameters in Table 11 were developed from literature and Adams optimizer.

Table 11. Autoencoder Model Parameter for Polymer Price forecast in Southeast Asia.

Parameter Name	Value
NN	GRU/GRU
Layer	1
Hidden	32
EPOCHS	50
Batch Size	64
Teacher Forcing	0.2
Learning Rate	0.025
Weight Decay	1.25
Batch Size	128

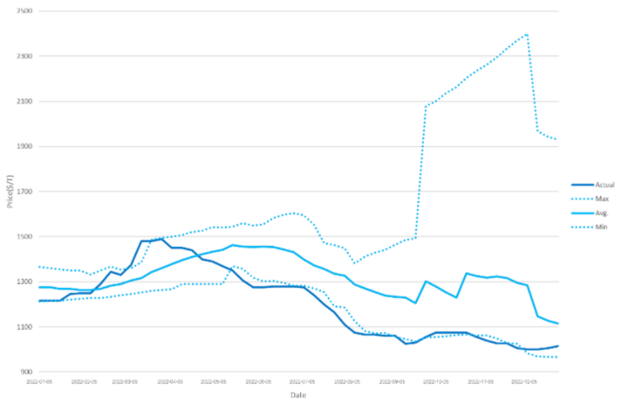

Fig. 17. Result of for Polymer Price forecast in Southeast Asia by Autoencoder.

The Autoencoder deep learning model is used to predict the HDPE product's pricing from January 2022 to December 2022. The model shows maximum, minimum, and average price forecasts, with dark grey representing the actual market price. The average forecast results are lowly correlated with the actual price, with MAE, MAPE and SMAPE values of 64, 24.9%, and 24.5% respectively. The average forecast exceeds the actual market price, indicating high levels of fluctuation in the delivered results.

d) TCNRNN

TCRNNs are hybrid neural networks which combine strengths for time series forecasting CNN and RNN. The model. The parameters in Table 12 and Table 13 were developed from literature and Adams optimizer.

Table 12. TCN Parameter for TCNRNN Model Parameter for Polymer Price Forecast in Southeast Asia.

Parameter Name	Value
TCN Layer	1
Filters	8
Kernel Size	2
Dilation Base	2
Dropout	0.2
Epochs	50
Learning Rate	0.025
Weight Decay	1.25
Batch Size	64

Table 13. RNN Parameter for TCNRNN Model Parameter for Polymer Price forecast in Southeast Asia.

Parameter Name	Value
Cell	gru
RNN Layer	1
Hidden Layer	16
Bidirectional	True
Epochs	50
Learning Rate	0.025
Weight Decay	1.25
Batch Size	64

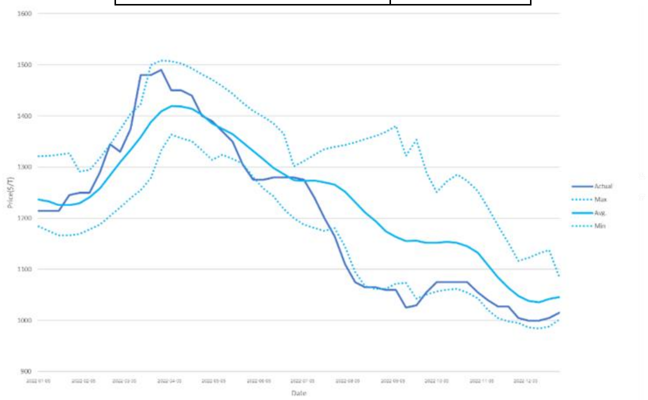

Fig. 18. Result of Polymer Price Forecast in Southeast Asia by TCNRNN.

Figure 18 shows projected HDPE product prices from January 2022 to December 2022 using the TCNRNN deep learning model. The dark blue dotted line represents the maximum, the light blue dotted line minimum, and the solid blue line average price forecasts. The dark grey colour corresponds to the current market value of HDPE. The MAE, MAPE and SMAPE values were 47,19.9% and 20.1% respectively. TCNRNN model shows a modest to high correlation between the price forecast and the actual market price.

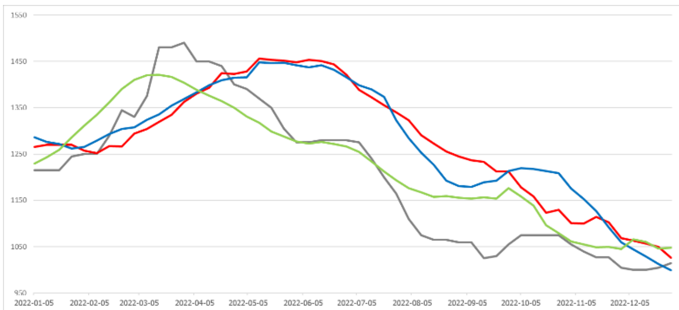

Fig. 19. Forecasted Results from Autoencoder, LSTM, and TCNRNNs.

Figure 19 shows the predicted HDPE product prices from January to December 2022, using deep learning models Autoencoder, LSTM, and TCNRNNs. The dark grey color represents the actual market price, with variations at different time intervals.

DOI:10.4186/ej.2024.28.12.79

Table 14. MAPE and SMAPE of Autoencoder, LSTM, and TCNRNN.

Error Rate	Autoencoder	LSTM	TCNRNN
Unit	%		
SMPE	24.5%	20.10%	20.10%
MAPE	24.9%	19.90%	19.90%
MAE	64	75	47

Table 14 compares MAPE, SMAPE and MAE for three deep learning models: Encoder-Decoder, Long Short-Term Memory (LSTM), and Temporal Convolutional Neural Network (TCNRNN). The TCNRNN model outperforms the other two in error measurement. All three models show a similar tendency to market price within a specified period but surpass market value. TCNRNN's forecasting tends to align with the actual market, while Autoencoder and LSTM models have low error rates.

The price forecast model was improved by incorporating unstructured data, particularly ICIS petrochemical news, which includes global market conditions and sentiment. Pre-trained text models were used, and the model could be executed by importing the library. The text encoder models were assessed under two scenarios, demonstrating their performance in the full report and specific section report scenarios.

Table 15. Comparing the Performance of Text Encoder Model for Full Report Scenario.

	MAPE	SMAPE	Time
Unit	(%)	(%)	(min/round)
Glove	53%	53%	243.74
FastText	59%	54%	275.34
NumberBatch	58%	59%	241.23
Sent2Vec	33%	34%	433.61
SentenceEncoder	39%	38%	392.12

Table 16. Comparing the Performance of Text Encoder Model for Specific Section Scenario.

	MAPE	SMAPE	Time
Unit	(%)	(%)	(min/round)
Glove	56%	57%	144.51
FastText	62%	58%	191.13
NumberBatch	59%	61%	146.14
Sent2Vec	37%	38%	231.29
SentenceEncoder	41%	41%	211.13

Transformer-based models, which use smaller vocabularies and require more time to run, offer a significant advantage in capturing intricate semantic information. These models use deeper and more sophisticated neural networks, making them prevalent in contemporary NLP applications. The SentenceEncoder architecture was chosen for its superior performance and satisfactory average runtime compared to other models, such as Sent2Vec, which outperformed other models but consumed more time and resources. Overall, Transformers are preferred for their ability to capture complex semantic information.

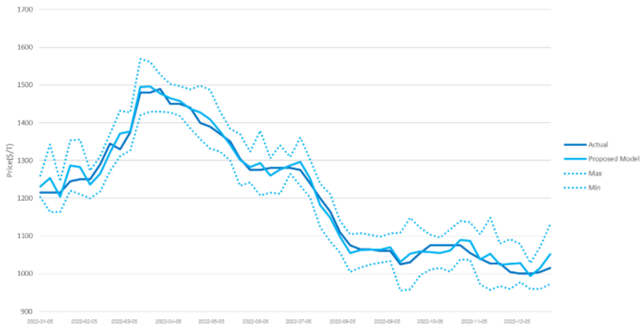

Fig. 20. Result of Polymer Price Forecast in Southeast Asia by Deep Learning with Sentiment Analysis.

According to results from Table 14 and Table 16.

Table 17. Error Rate of Polymer Price Forecast in Southeast Asia by TCNRNN with NLP.

	MAE	MAPE	SMAPE
Error Rate	38	17.1%	18.2%

Figure 20 shows the projected HDPE product price from January 2022 to December 2022, using TCNRNN and SentenceEncoder deep learning models. The upper and lower limits are indicated by dark blue dotted lines, while the average price forecast is solid blue. The MAE, MAPE and SMAPE values are 38, 17.1%, and 18.2%, indicating a high correlation between the price forecast and the actual market price.

ANOVA is a statistical method used to compare the mean performance of various machine learning algorithms, with a significant deviation from the overall mean indicating performance differences, and a p-value below 0.05 indicating substantial differences [92].

The study tested hypotheses with a significance level of 0.05. ANOVA results showed a price forecast error rate of 43% for LSTM EncoderDecoder, TCNRNN, and TCNRNN with Text Model. The data showed a total Sum of Square (SS) of 0.8797, a total Degree of Freedom (df) of 207, a Mean Square (MS) of 0.2446 between groups, and an MS within groups of 0.0007. The F-value was 342.139, indicating greater between-group variability than within-group variability. The P-value was below the significance level of $\alpha = 0.05$, indicating a statistically significant difference between the means of the four groups.

Table 18. Statistical Test Summary.

Groups	Count	Sum	Average	Variance
LSTM	52	22.7266	0.43705	0.0001984
Autoencoder	52	26.513	0.5098653	0.0014355
TCNRNN	52	21.1595	0.4069134	0.0003164
TCNRNN with NLP	52	17.3393	0.3334480	0.0009093

Source	SS	df	MS	F	P-value	F crit
Between Groups	0.7338580	3	0.2446193	342.13996	2.607E-79	2.6488
Within Groups	0.1458535	204	0.0007149		(p<0.05)	
Total	0.8797116	207				

5.6. Business Process Adjustment

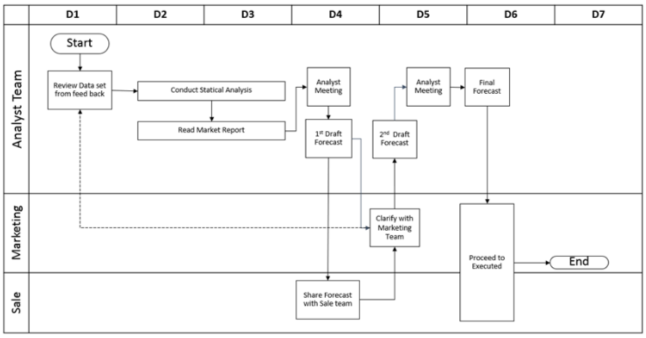

Fig. 21. Flow Chart of Current Forecast Process.

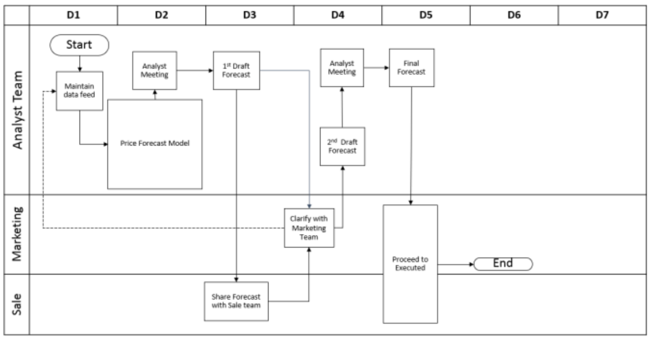

Fig. 22. Flow Chart of Proposed Forecast Process.

The current forecast process, as illustrated in Fig. 21, is compared with the proposed forecast process depicted in Fig. 22. The existing forecasting process is labor-intensive and time-consuming, encompassing data preparation, analysis, model adjustments, and iterations. It relies significantly on subjective judgment, which can cause conflicts between marketing and sales teams. The process is hindered by unorganized data and lacks the capability for real-time data and sentiment analysis. Furthermore, it necessitates manual communication among data science, marketing, and sales teams due to the absence of an interface for forecast exploration. This method is imprecise, biased, and requires persuasion for implementation. Forecasts are generated weekly, projecting three months ahead.

Conversely, the proposed procedure enhances automation and reduces the need for team discussions. It employs automated data ingestion and modeling techniques, minimizing iterations and saving time. By leveraging large volumes of both structured and unstructured data, it improves accuracy. The process also reduces manual communication among analysts, marketing, and sales teams by providing an interface for accessing projections.

5.7. Forecast Model Application

Table 19. Example of Weekly Market price.

	Week1	Week2	Week3	Week4	Week5
Market	1,005	1,000	1,015	1,020	1,000
Change	0	-5	+15	+5	-20

This disclaimer clarifies that the example volume and market prices provided are randomly generated to illustrate the application of a price forecast model.

As shown in Table 19, the market price initially decreased from \$1005/T to \$1000/T (-\$5/T) between week 1 and week 2. It then rose from \$1000/T to \$1015/T (+\$15/T) from week 2 to week 3. Subsequently, the price increased again from \$1015/T to \$1020/T (+\$5/T) between week 3 and week 4. Finally, the market price declined from \$1020/T to \$1000/T (-\$20/T) from week 4 to week 5.

This paper recommends employing a price forecast model to capitalize on market price volatility in the following two scenarios.

Scenario 1: Price uptrend

Upon the model forecasting a positive upward trend, the sales and marketing team should halt the product presentation process. They should anticipate a price increase before resuming negotiations.

Scenario 2: Price downtrend

When the model forecasts a negative, declining trend, the sales and marketing team should accelerate the quotation process and broaden the range of products offered to the customer.

6. Conclusion

This study focuses on the petrochemical industry and its impact on market prices. It uses economic and pricing theory to analyze a company case study focusing on the commodity price of polymer products in the Southeast Asian market, focusing on Oil/Gas and its derivatives and macro-economic indicators. The research uses various advanced libraries such as NumPy, SciPy, Pandas, TensorFlow, and Natural Language Toolkit (NLTK) to analyze the data.

Feature engineering is crucial for developing deep learning models for forecasting, as it significantly enhances the performance and accuracy of machine learning models by identifying hidden patterns within data. Accurate data from users is essential for machine algorithms to function effectively. The research proposes a deep learning model suitable for the petrochemical industry, but requires data cleaning for missing value imputation and identifier outliers, and revisiting input features using recommended procedures like the Stationary test, Correlation test, and features important test.

The study explores several deep learning model architectures for analyzing both structured and unstructured data, with the goal of predicting market prices for High-Density Polyethylene (HDPE) commodities. Three deep learning model architectures were identified as the most effective and widely utilized for predicting time series prices: LSTM, EncoderDecoder, and TCNRNN. All three models indicate that the forecasted results for the first quarter of 2022 have a lower error rate compared to the other quarters of the year.

The price of HDPE was influenced by strong market mood, volatility, and unfavourable circumstances, including geopolitical situation, movement of oil and gas party or OPEC's actions, and global logistical/freight situation. The price forecast model was improved by incorporating unstructured data, particularly ICIS petrochemical news, which includes global market conditions and sentiment. Pre-trained models were imported for execution, enhancing the model's capabilities.

A comparison of five text deep learning models, FastText, GloVe, NumberBatch, Sent2Vec, and Universal Sentence Encoder, will be conducted to determine the most suitable model for a business. Transformer models outperform classical models, even with smaller versions and no customized text cleaning procedures. The SentenceEncoder architecture was chosen due to its superior performance and satisfactory average runtime compared to the standard, despite the significant time resource consumption of Sent2Vec.

Table 20. Comparison between business in-use model and proposed model.

	MAE	MAPE	SMAPE
In use Model	184	45.10%	44.10%
Proposed Model	38	17.10%	18.20%
Improvement	79%	62%	59%

Table 20 presents a comparison between a currently utilized business model, which relies on linear computation and human reasoning, and a proposed deep learning model. The study revealed that the deep learning model significantly reduced the error rate. Specifically, the MAE, MAPE and SMAPE of the proposed model improved by 79%, 62%, and 59%, respectively, compared to the existing model.

The findings suggest that a combination of the TCNRNN deep learning model and the Sentence Encoder deep learning model are the most appropriate ways for predicting petrochemical prices. By integrating both structured and unstructured deep learning models, the accuracy of the model and understanding of important factors affecting pricing in a volatile, unpredictable, complex, and ambiguous (VUCA) global environment is improved.

6.1. Academic Contribution

This research presents an Ensemble model of Deep learning architecture that can predict polymer prices in the Southeast Asia market using both structured and unstructured data. It uses feature engineering to identify key features with high correlations with the price. The prediction timeframe is weekly, with updates updated every week. This is the first time commodity price forecasting using deep learning has been conducted within this specific time frame, according to the literature review.

6.2. Practical contribution

This research explores the impact of feature engineering on polymer price forecasting, highlighting its potential benefits for both deep-learning models and human analysts. It examines the current implementation of corporate case study business practices, identifying challenges and disadvantages, and proposing alternative approaches to improve outcomes. The analysis can be applied to other businesses in the same or different industry sectors.

6.3. Further research

The research focuses on predicting the market price of HDPE in the Southeast Asia region using Deep Learning and NLP models. The top-performing Deep learning model, TCNRNN, is combined with all five NLP models to achieve the optimal combination. The study suggests conducting additional experiments to address resource constraints. The framework for forecasting is already in place, and the prediction target can be changed by adjusting feature input and rerunning the procedure. However, most components are still offline and rely on manual processes, requiring human intervention. The study improves efficiency by utilizing cloud computing architecture, automated scripts, and APIs, enhancing the model workflow. The goal is to improve the efficiency of the model workflow. In addition, the concepts in this research can also be extended to other applications, for example, as discussed in [92], [93] and [94].

References

[1] S. G. Platt, "Polyethylene resins linear high-density HDPE," S&P Global Platt, 2022.

[2] S. G. Platt, "Polyethylene resins low-density LDPE," S&P Global Platt, 2022.

[3] S. G. Platt, "Polyethylene resins low-density LLDPE," S&P Global Platt, 2022.

[4] M. E. Porter, "The five competitive forces that shape strategy," *Harvard Business Review*, vol. 86, no. 1, 2008.

[5] G. Armstrong, S. Adam, S. Denize, and P. Kotler, *Principles of Marketing*. Pearson Australia, 2014.
[6] N. G. Mankiw, *Principles of Economics*. Cengage Learning, 2021.
[7] N. Frumkin, *Guide to Economic Indicators*. Routledge, 2015.
[8] T. Callen, "Gross Domestic Product (GDP): An Economy's All." [Online]. Available: https://www.imf.org/external/pubs/ft/fandd/basics/14_gdp.htm#:~:text=GDP%20is%20composed%20of%20goods,the%20residents%20of%20a%20country
[9] S&P Global Market Intelligence. "PMI - Purchasing Managers' Index." [Online]. Available: https://www.spglobal.com/market-intelligence/en/solutions/products/pmi
[10] L. J. Christiano, "A survey of measures of capacity utilization (Mesures de l'utilisation des capacités de production) (Estudio de mediciones de utilización de la capacidad)," Staff Papers - International Monetary Fund, pp. 144–198, 1981.
[11] IBM, "Structured vs. Unstructured Data." [Online]. Available: https://www.ibm.com/blog/structured-vs-unstructured-data
[12] B. Clarke, E. Fokoue, and H. H. Zhang, *Principles and Theory for Data Mining and Machine Learning*. Springer Science & Business Media, 2009.
[13] W. McKinney, *Python for Data Analysis*. O'Reilly Media, Inc., 2022.
[14] B. Khan, S. Jan, W. Khan, and M. I. Chughtai, "An overview of ETL techniques, tools, processes and evaluations in data warehousing," *Journal on Big Data*, vol. 6, 2024.
[15] J. Caserta and R. Kimball, *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley, 2013.
[16] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. New York: Springer, 2009.
[17] J. M. Kanter and K. Veeramachaneni, "Deep feature synthesis: Towards automating data science endeavors," in *International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2015, pp. 1–10.
[18] T. Qiu and Y. Guo, *Signal Processing and Data Analysis*. Walter de Gruyter GmbH & Co. KG, 2018.
[19] T. Y. Wang and H. M. Chiang, "Solving multi-label text categorization problem using support vector machine approach with membership function," *Neurocomputing*, vol. 74, no. 17, pp. 3682–3689, 2011.
[20] G. Appel, *Technical Analysis: Power Tools for Active Investors*. FT Press, 2005.
[21] J. Bollinger, *Bollinger on Bollinger Bands*. New York: McGraw-Hill, 2002.
[22] K. K. Nicodemus and J. D. Malley, "Predictor correlation impacts machine learning algorithms: Implications for genomic studies," *Bioinformatics*, vol. 25, no. 15, pp. 1884–1890, 2009.
[23] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.
[24] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.
[25] S. Johansen, "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, vol. 12, no. 2–3, pp. 231–254, 1988.
[26] S. Johansen, "Cointegration: An overview," Department of Applied Mathematics and Statistics, University of Copenhagen, 2004.
[27] S. Johansen, "Cointegration: Overview and development," in *Handbook of Financial Time Series*. 2009, pp. 671–693.
[28] M. Wojtas and K. Chen, "Feature importance ranking for deep learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 5105–5114, 2020.
[29] XGBoost Developers, "XGBoost Documentation." [Online]. Available: https://xgboost.readthedocs.io/en/stable/
[30] G. Montavon, W. Samek, and K. R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, pp. 1–15, 2018.
[31] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems with Applications*, vol. 184, p. 115537, 2021.
[32] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999-7019. 2021.
[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[34] S. Liu, C. Zhang, and J. Ma, "CNN-LSTM neural network model for quantitative strategy analysis in stock markets," in *Neural Information Processing: 24th International Conference, ICONIP 2017*, Guangzhou, China, Proceedings, Part II, Springer International Publishing, Nov. 14–18, 2017, pp. 198–206.
[35] A. Roy Choudhury, S. Abrishami, M. Turek, and P. Kumar, "Enhancing profit from stock transactions using neural networks," AI Communications, vol. 33, no. 2, pp. 75–92, 2020.
[36] W. Li and D. M. Becker, "Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling," *Energy*, vol. 237, p. 121543, 2021.
[37] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[38] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *The Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[39] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An open multilingual graph of general knowledge," in *Proc. AAAI Conference on Artificial Intelligence*, 2017, vol. 31, no. 1.

[40] M. Pagliardini, P. Gupta, and M. Jaggi, "Unsupervised learning of sentence embeddings using compositional n-gram features," 2017, *arXiv:1703.02507*.

[41] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil, "Universal sentence encoder," 2018. [Online]. Available: https://arxiv.org/pdf/1803.11175.pdf

[42] N. Amral, C. S. Ozveren, and D. King, "Short term load forecasting using Multiple Linear Regression," in *2007 42nd International Universities Power Engineering Conference*, 4–6 Sept. 2007, pp. 1192–1198, doi: 10.1109/UPEC.2007.4469121.

[43] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques – Part II: Soft computing methods," Expert Systems with Applications, vol. 36, no. 3, Part 2, pp. 5932–5941, Apr. 2009, doi: 10.1016/j.eswa.2008.07.006.

[44] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning, JMLR Workshop and Conference Proceedings*,2012, pp. 37–49.

[45] M. Bildirici and Ö. Ö. Ersin, "Nonlinearity, volatility and fractional integration in daily oil prices: Smooth transition autoregressive ST–FI (AP) GARCH models," *Romanian Journal of Economic Forecasting*, vol. 3, pp. 108–135, 2014.

[46] S. Boubaker, Z. Liu, and Y. Zhang, "Forecasting oil commodity spot price in a data-rich environment," *Annals of Operations Research*, pp. 1–18, 2022.

[47] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting," *IEEE Access*, vol. 8, pp. 71326–71338, 2020.

[48] J. F. Chen, W. L. Chen, C. P. Huang, S. H. Huang, and A. P. Chen, "Financial time-series data analysis using deep convolutional neural networks," in 2016 *7th International Conference on Cloud Computing and Big Data (CCBD)*, 16–18 Nov. 2016, pp. 87–92, doi: 10.1109/CCBD.2016.027.

[49] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[50] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[51] J. Doering, M. Fairbank, and S. Markose, "Convolutional neural networks applied to high-frequency market microstructure forecasting," in *2017 9th Computer Science and Electronic Engineering (CEEC)*, 27-29 Sept. 2017, pp. 31-36, doi: 10.1109/CEEC.2017.8101595.

[52] G. Elliott and A. Timmermann, "Economic forecasting," *Journal of Economic Literature*, vol. 46, no. 1, pp. 3-56, 2008. [Online]. Available: http://www.jstor.org/stable/27646946

[53] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654-669, 2018

[54] C. García-Martos, J. Rodríguez, and M. J. Sánchez, "Modelling and forecasting fossil fuels, CO_2 and electricity prices and their volatilities," *Applied Energy*, vol. 101, pp. 363-375, 2013

[55] L. P. Hansen, "Time-series econometrics in macroeconomics and finance," *Journal of Political Economy*, vol. 125, no. 6, pp. 1774-1782, 2017.

[56] R. Hrasko, A. G. C. Pacheco, and R. A. Krohling, "Time series prediction using restricted Boltzmann machines and backpropagation," *Procedia Computer Science*, vol. 55, pp. 990-999, 2015.

[57] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems with Applications*, vol. 184, p. 115537, 2021.

[58] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, vol. 10, no. 3, pp. 215-236, 1996.

[59] A. Kanwal, M. F. Lau, S. P. Ng, K. Y. Sim, and S. Chandrasekaran, "BiCuDNNLSTM-1dCNN—A hybrid deep learning-based predictive model for stock price prediction," *Expert Systems with Applications*, vol. 202, p. 117123, 2022.

[60] N. Kohzadi, M. S. Boyd, B. Kermanshahi, and I. Kaastra, "A comparison of artificial neural network and time series models for forecasting commodity prices," *Neurocomputing*, vol. 10, no. 2, pp. 169-181, 1996

[61] W. Kristjanpoller and M. C. Minutolo, "Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7245-7251, 2015

[62] T. T. H. Le and T. N. Nguyen, "A hybrid model based on ARIMA and artificial neural network to forecast consumer price index: The case of Vietnam," in *Optimal Transport Statistics for Economics and Related Topics*, N. Ngoc Thach, V. Kreinovich, D. T. Ha, and N. D. Trung Eds. Cham: Springer Nature Switzerland, 2024, pp. 449-464.

[63] A. Lendasse, E. de Bodt, V. Wertz, and M. Verleysen, "Non-linear financial time series forecasting-Application to the Bel 20 stock market index," European Journal of Economic and Social Systems, vol. 14, no. 1, pp. 81-91, 2000.

[64] P. Li, C. Jing, T. Liang, M. Liu, Z. Chen, and L. Guo, "Autoregressive moving average modeling in the financial sector," in *2015 2nd International Conference on*

Information Technology, Computer, and Electrical Engineering (ICITACEE), IEEE, 2015, pp. 68-71.

[65] X. Li, W. Shang, and S. Wang, "Text-based crude oil price forecasting: A deep learning approach," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1548-1560, 2019.

[66] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999-7019, 2021.

[67] W. C. Lin, C.-F. Tsai, and J. R. Zhong, "Deep learning for missing value imputation of continuous data and the effect of data discretization," *Knowledge-Based Systems*, vol. 239, p. 108079, 2022, doi: https://doi.org/10.1016/j.knosys.2021.108079.

[68] S. Liu, C. Zhang, and J. Ma, *CNN-LSTM Neural Network Model for Quantitative Strategy Analysis in Stock Markets*. 2017, pp. 198-206.

[69] J. Peng, Z. Li, and B. Drakeford, "Dynamic characteristics of crude oil price fluctuation-from the perspective of crude oil price influence mechanism," *Energies*, vol. 13, 08/29 2020, doi: 10.3390/en13174465.

[70] J. Shi,, M. Jain,, & G. Narasimhan, "Time series forecasting (TSF) using various deep learning models," 2022, *arXiv:2204.11115*.

[71] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[72] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and Bi-directional LSTM model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 24-26 Oct. 2020, pp. 87-92, doi: 10.1109/NILES50944.2020.9257950.

[73] M. G. C. Tapia and C. A. C. Coello, "Applications of multi-objective evolutionary algorithms in economics and finance: A survey," in *2007 IEEE Congress on Evolutionary Computation*, 25-28 Sept. 2007

[74] R. van Eyden, M. Difeto, R. Gupta, and M. E. Wohar, "Oil price volatility and economic growth: Evidence from advanced economies using more than a century's data," *Applied Energy*, vol. 233-234, pp. 612-621, 2019

[75] P. H. Vuong, T. T. Dat, T. K. Mai, and P. H. Uyen, "Stock-price forecasting based on XGBoost and LSTM," *Computer Systems Science & Engineering*, vol. 40, no. 1, 2022.

[76] T.-Y. Wang and H.-M. Chiang, "Solving multi-label text categorization problem using support vector machine approach with membership function," *Neurocomputing*, vol. 74, no. 17, pp. 3682-3689, 2011

[77] L. Wei and Z. Zhen-gang, "Based on time sequence of ARIMA model in the application of short-term electricity load forecasting," in *2009 International Conference on Research Challenges in Computer Science*, 28-29 Dec. 2009, pp. 11-14

[78] K. A. Weldon, J. Ngechu, N. Everlyne, N. Njambi, and K. Gikunda, "Petroleum prices prediction using data mining techniques—A Review," 2022, *arXiv:2211.12964*.

[79] L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting," *Applied Energy*, vol. 167, pp. 135-153, 2016.

[80] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu, "Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting," *Applied Energy*, vol. 180, pp. 213-233, 2016.

[81] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Applied Energy*, vol. 134, pp. 102-113, 2014.

[82] Z. Zhao, C. Fu, C. Wang, and C. J. Miller, "Improvement to the prediction of fuel cost distributions using ARIMA Model," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 5-10 Aug. 2018, pp. 1-5

[83] R. J. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, vol. 793. John Wiley & Sons 2019.

[84] R. Medar, V. S. Rajpurohit and B. Rashmi, "Impact of training and testing data splits on accuracy of time series forecasting in machine learning," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, Pune, India, 2017, pp. 1-6

[85] M. Hasanov, M. Wolter and E. Glende, "Time series data splitting for short-term load forecasting," in *PESS + PELSS 2022; Power and Energy Student Summit*, Kassel, Germany, 2022, pp. 1-6.

[86] C. Desai, "Comparative analysis of optimizers in deep neural networks," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 10, pp. 959-962, 2020.

[87] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, "The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study," *Multimedia Tools and Applications*, vol. 82, no. 11, 16591-16633, 2023.

[88] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference for Learning Representations*. 2015.

[89] N. Mehdiyev, D. Enke, P. Fettke, and P. Loos. "Evaluating forecasting methods by considering different accuracy measures," *Procedia Computer Science*, vol. 95, pp. 264-271, 2016.

[90] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, no. 1, pp. 69-80, 1992

[91] J. Zhao, Z. Dalin, S. Liang, H. Kang, and Q. Liu, "Prediction model for stock price trend based on

recurrent neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 745-753, Jan. 2021.

[92] T. Arnonwattana and P. Chutima, "Leveraging partner country factors in deep learning for Thailand's forecasted inflation accuracy enhancement," *Eng. J.*, vol. 28, no. 6, pp. 37-58, Jun. 2024.

[93] W. Khrakhuean and P. Chutima, "Real-time induction motor health index prediction in a petrochemical plant using machine learning," *Eng. J.*, vol. 26, no. 5, pp. 91-107, May 2022.

[94] W. Cheevachaipimol, B. Teinwan, and P. Chutima, "Flight delay prediction using a hybrid deep learning method," *Eng. J.*, vol. 25, no. 8, pp. 99-112, Aug. 2021.

Narongchon Amapat received a Bachelor of Engineering from King Mongkut's University of Technology Thonburi (KMUTT), a master's degree from Chulalongkorn University and King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand. He is a full-time Analyst at PTT Global Chemical PCL, Thailand. His scope of work and research is the downstream of the petrochemical business and optimisation in operations management by advanced analytic model.

Professor Parames Chutima received a Bachelor of Engineering from Chulalongkorn University, Master's degrees from Chulalongkorn University and Asian Institute of Technology, and a PhD in Manufacturing Engineering and Operations Management from the University of Nottingham, UK. He is a full Professor in the Faculty of Engineering at Chulalongkorn University, Thailand, and the Director of the Regional Centre for Manufacturing Systems Engineering, Chulalongkorn University, Thailand. His research interests include multi-objective optimisation in operations management, production planning and control of assembly lines, just-in-time production systems and simulation modelling. He is the author and co-author of many books and scientific articles in prestigious international journals and conferences.