Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

Shehar Yaar Aziz
Ole Jacob Simensen

Forecasting German Natural Gas Prices using Temporal Fusion Transformer

Master's thesis in Industrial Economics and Technology Management
Supervisor: Einar Belsom
December 2023

NTNU
Norwegian University of Science and Technology

Shehar Yaar Aziz
Ole Jacob Simensen

Forecasting German Natural Gas Prices using Temporal Fusion Transformer

Master's thesis in Industrial Economics and Technology Management
Supervisor: Einar Belsom
December 2023

Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

NTNU
Norwegian University of
Science and Technology

Preface

This Master of Science thesis was conducted at the Norwegian University of Science and Technology within the Department of Industrial Economics and Technology Management during the fall semester of 2023. It serves as a representation of the practical application of our academic knowledge, specifically at the intersection of technology and economics. Our intent is for this work to offer valuable insights, contributing to an understanding of the dynamics between deep learning and commodity market behaviors.

We thank our supervisor, Einar Belsom, for his valuable guidance throughout the course of this research project. His advice on relevant literature, continuous feedback, and knowledge on how to apply deep learning to commodity markets has been beneficial. The deep learning model in this thesis could only have been developed with his eagerness to help and insights.

Abstract

We utilize Temporal Fusion Transformer (TFT), a sophisticated deep learning model, to forecast German natural gas prices spanning from January 2012 to February 2023. Our main objective is to evaluate the effectiveness of TFT as a forecasting tool, specifically for German natural gas prices, with a 14-day forecasting horizon. Our study extends current research by employing the relatively new TFT model. In addition to its proven effectiveness in terms of performance, the TFT model also offers in-depth interpretability, enabling the analysis of pattern detection.

Among a broad selection of features relevant to natural gas prediction, we identify an effective combination for the TFT model that results in accurate price predictions. This is accomplished by leveraging the transparency of the TFT model, where we filter out potential noisy features. Some of the most prominent features include other trading hub prices, temperature data, oil volatility and Google searches for the term "war".

To evaluate the robustness of the forecasts, we partition the time series into distinct periods characterized by different price behavior. We find that the TFT model for many prediction periods, significantly outperforms benchmark models in terms of Mean squared error (MSE). Additionally, TFT excels in accurately forecasting the direction of price movements, even during periods of heightened market volatility.

Sammendrag

Vi benytter Temporal Fusion Transformer (TFT), en avansert dyp læremodell, for å predikere tyske naturgasspriser fra januar 2012 til februar 2023. Hovedmålet vårt er å evaluere effektiviteten til TFT som et prognoseverktøy, spesielt for tyske naturgasspriser, med en prediksjonshorisont på 14 dager. Oppgaven utvider eksisterende forskning ved å ta i bruk den relativt nye TFT-modellen. I tillegg til at tidligere studier har trukket frem gode prestasjoner, har TFT-modellen også grundig tolkningsmulighet, som muliggjør analyse av mønstergjenkjenningen i modellen.

Blant et bredt utvalg input variable som er relevante for naturgassprediksjon, identifiserer vi en effektiv kombinasjon for TFT-modellen som resulterer i nøyaktige prediksjoner. Dette oppnås ved å utnytte TFT-modellens transparens, der vi filtrerer ut variable som ser ut til å tilføre støy. Noen av de mest fremtredende variablene er; priser fra andre gass huber, temperaturdata, oljevolatilitet og Google-søk etter ordet "krig".

For å vurdere robustheten til prediksjonene deler vi tidsserien inn i distinkte perioder basert på prisatferd. Resultatene avdekker at TFT-modellen, i mange test perioder, overgår referansemodeller signifikant når det gjelder gjennomsnittlig kvadratfeil. I tillegg, utmerker TFT seg ved å nøyaktig forutsi retningen av prisbevegelser, selv i perioder med økt markedsvolatilitet.

Table of Contents

List of Figures

List of Tables

Abbreviations

ANN	Artificial neural network
ARCH	Autoregressive conditional heteroskedasticity
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
ASVI	Abnormal search volume indexes
BDI	Baltic dry index
CAPM	Capital asset pricing model
CO_2	Carbon dioxide
EMH	Efficient market hypothesis
GBM	Gradient boosting machine
GARCH	Generalized autoregressive conditional heteroskedasticity
GPL	GASPOOL
GRN	Gated residual networks
LNG	Liquid natural gas
LSTM	Long short-term memory
MAE	Mean absolute error
MSE	Mean squared error
MTPA	Million tonnes per annum
MT	Million tonnes
MMbtu	Metric million British thermal units
NBP	National Balancing Point
NCG	NetConnect Germany
PMI	Purchasing managers index
RNN	Recurrent neural networks
RMSE	Root mean squared error
SVM	Support vector machine
TFT	Temporal Fusion Transformer
TTF	Title Transfer Facility
VSN	Variable selection network
XGBoost	eXtreme Gradient Boosting

Chapter 1

Introduction

In recent years, the German natural gas market has experienced substantial transformations. The energy landscape in Europe underwent considerable upheaval, triggered by the Russian assault on Ukraine and the subsequent suspension of Russian natural gas deliveries to Europe [Korosteleva, 2022]. The disruption was further intensified by the sabotage of Nord Stream, the pipeline connecting Russia to Germany through the Baltic Sea [Jacobsen & Abnett, 2022]. This has led to significant shifts in the composition of Germany's gas imports [Eckert & Steitz, 2023]. Historically reliant on Russia as its primary source, Germany faced an abrupt cessation of gas deliveries. Consequently, the country had to implement costly emergency measures to secure alternative suppliers from the global market [Kemfert et al., 2022]. In response to these challenges, Germany has proactively taken numerous measures to diversify its gas supply sources [Eckert & Steitz, 2023]. There is a particular emphasis on expediting the development of liquefied natural gas (LNG) import infrastructure, a move characterized by an unprecedented and rapid pace [Kemfert et al., 2022].

Amidst the recent substantial transformations in the German natural gas market and the broader European energy landscape, stakeholders meet enhanced needs of accurately predicting natural gas prices. This prompts the question: Can deep learning be a viable tool for predicting future gas prices in this evolving landscape?

On natural gas price forecasting, recent studies have explored the efficacy of advanced machine learning models. In the study by Ali et al. [2021], a deep neural network (DNN) is employed for forecasting natural gas prices for the American natural gas trading hub, Henry Hub. The research utilizes an extensive dataset spanning 281 months to develop and assess the predictive capabilities of the proposed DNN model. Comparative analysis against five contemporary machine learning models reveals that the suggested DNN model consistently outperforms its counterparts, showcasing superior performance across multiple evaluation metrics, including mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination. Similarly, Su et al. [2019] assesses the predictive capabilities of various machine learning models for Henry Hub monthly prices from 2001 to 2018. They evaluated the performance of artificial neural network (ANN), Support vector machine (SVM), Gradient Boosting Machines (GBM), and Gaussian Process Regression (GPR) to determine which model exhibited the highest level of accuracy. They used features such as crude oil prices, heating oil prices, temperature, natural gas supply and demand, imports, and storage levels to predict future prices. Their findings indicated that the ANN models outperformed other models significantly in this endeavor.

Additionally, Ram et al. [2019] utilizes a neural network model to forecast prices at five major natural gas exchange centers in Europe. With an R^2 performance evaluation index of 98%, indicating the well-fitted nature of the neural network model to the provided data series, the model demonstrates remarkable accuracy in forecasting natural gas prices. The study's findings underscore the efficacy of the ANN method, revealing its ability to predict gas prices in European gas hubs with high precision.

Concerning forecasts specific to German natural gas, the research conducted by Busse et al. [2012] is comprehensive. Their research examined various features gathered through expert interviews and assessed the influence of individual feature sets when utilizing a NARX Neural Network for price movement forecasts. Their results suggest that implementing a machine learning model to predict German gas prices could lead to accurate performance. Additionally, they identified an optimal feature set, including weather data, exchange rates, and prices of major hubs.

Lim et al. [2021] introduced a groundbreaking architecture known as the Temporal Fusion Transformer (TFT). This model combines advanced multi-horizon forecasting capabilities with the ability to unravel the temporal dynamics of the underlying data. Comprising three key components encoder, decoder, and recurrent layers. The TFT model processes input features using self-attention mechanisms to derive meaningful representations. The decoder then utilizes these encoded representations to generate predictions for the target variable. The recurrent layers capture temporal dependencies by incorporating previous predictions as inputs, enhancing the model's ability to understand and adapt to evolving patterns [Wang et al., 2022]. Additionally, when compared against competing methods like ARIMA and DeepAR, as detailed in Lim et al. [2021], TFT consistently outperforms them, exhibiting performance improvements ranging from 3% to 26%.

To our knowledge, no paper has investigated the effectiveness of the TFT model in gas price predictions. The field of predicting German natural gas prices using machine learning also remains an area where limited extensive research has been conducted. By utilizing a TFT model incorporating a comprehensive feature set we aim to bridge the gap in the existing literature and extend the insights from Busse et al. [2012]. Following applying the TFT model, we conduct an in-depth post-model analysis. This involves exploring the importance of features and identifying significant events within the dataset. This multifaceted analysis aims to provide a nuanced understanding of the factors influencing German natural gas price dynamics. Lastly, to enrich our comparative insights, we utilize Autoregressive Integrated Moving Average (ARIMA) and eXtreme Gradient Boost (XGBoost) as benchmark models. ARIMA, a widely utilized time series forecasting method, is known for its effectiveness in capturing linear trends and seasonal patterns. Furthermore, XGBoost, a powerful ensemble learning technique, has gained popularity for its ability to handle complex relationships within data and achieve robust predictive performance [NVIDIA, 2023]. The results obtained from all three models are compared, allowing for a comprehensive evaluation of their respective forecasting capabilities. This comparative approach adds a valuable layer to our study, contributing insights into the relative strengths and weaknesses of TFT, ARIMA, and XGBoost models in the context of gas price predictions.

The remainder of this study is structured as follows. We first present the relevant background information in Chapter 2. Chapter 3 provides a guide trough the evolution of econometrics. Then, an extensive description of the data is presented in Chapter 4, before the methodology is presented in Chapter 5. Results and the following discussion are presented in Chapter 6. Finally, concluding remarks are presented in Chapter 7.

Chapter 2

Background

Natural gas is a vital component of the modern energy landscape, with a significant presence worldwide in industrial, residential, and commercial sectors. This chapter aims to explain its multifaceted attributes, from its extraction to market dynamics, with a particular focus on the context of Germany. This is done to provide the reader with a more comprehensive understanding of the behavior of natural gas prices, as well as to understand the motivation behind incorporating multiple input features. First, a foundational understanding of natural gas is provided, covering its origin, extraction methodologies, and transportation process. Second, the supply and demand dynamics and the intricate pricing mechanism will be explained. Finally, the market and price considerations of natural gas in Germany will be analyzed.

2.1 Natural gas as a Fuel

According to Speight [2018, p. 6-10], natural gas, an odorless gaseous blend of hydrocarbons dominated by methane, may also contain trace elements of ethane, propane, butane, and other gases, as well as non-hydrocarbon gases like carbon dioxide and nitrogen. Widely recognized as a vital primary energy source globally, unprocessed natural gas—raw natural gas—constitutes approximately 20%

Furthermore, as highlighted by Speight [2018, p. 166-178], natural gas is efficiently conveyed through a network of pipelines, with the most extensive ones, known as transmission pipelines, often crossing provincial boundaries. Upon reaching its destination, local distribution companies or gas utilities distribute natural gas to residences and businesses. Additionally, natural gas can be stored in underground facilities for future utilization. An alternative method involves liquefaction, transforming natural gas into liquefied natural gas (LNG). Achieved by cooling the gas to approximately -160 degrees, LNG serves as a liquid state for natural gas. Its primary purpose is transporting natural gas to diverse markets. Upon arrival, LNG undergoes regasification and is distributed as pipeline natural gas. Liquefaction offers the advantage of reduced volume compared to natural gas, facilitating storage and transportation. In this state, the gas is transported via ships or specialized vehicles. Shipping across the sea necessitates using dedicated refrigerated vessels to maintain the gas's conditions.

As outlined by Speight [2018, p. 4-6], adopting natural gas represents a relatively recent development in energy usage. In the 20th century, coal and oil were the prevailing energy sources. However, the world faced an oil shortage in the 1970s, prompting concerns about resource availability. Before this crisis, the practical applications of natural gas were constrained. It was only during this period of scarcity that novel uses for natural gas emerged. The rapid expansion of its utilization ensued once natural gas transportation over extended distances became viable.

Speight [2018, p. 149-179] highlights that natural gas serves diverse purposes, including heating residential homes, powering various industrial processes, and fueling transportation. However, its most widespread application is in electricity generation. Natural gas power plants leverage gas turbines to produce electricity. Notably, these plants are renowned for their cost-effectiveness, standing out as some of the most economically viable facilities for power generation. Additionally, they offer enhanced flexibility compared to coal-fired plants, allowing swift adjustments in power output. Historically, in the United States, natural gas-fired plants were predominantly utilized during periods of high demand.

Speight [2018, p. 367-378] asserts that fossil fuels, including coal, crude oil, and natural gas, release pollutants into the atmosphere during the energy generation. However, natural gas, predominantly composed of methane, is the most environmentally friendly among these fuels. Its combustion yields a cleaner burn than other fossil fuels, resulting in fewer sulfur, carbon, and nitrogen emissions when contrasted with coal or oil. Additionally, the combustion of natural gas leaves minimal residual ash particles. These characteristics are anticipated to further drive the growing reliance on natural gas in the energy sector.

2.2 Supply and Demand Dynamics

Natural gas supply is driven by a combination of factors, including the demand for energy, economic considerations, environmental concerns, government policies, accessibility, and production costs. The production of natural gas has been increasing due to its lower carbon emissions compared to other fossil fuels [Balcombe et al., 2018]. As the drive to reduce climate change intensifies, natural gas is considered more environmentally friendly.

When modeling gas supply, Crow et al. [2018] places emphasis on specific technical variables connected to natural gas production. First, the magnitude of accessible natural gas reserves serves as a limiting factor, determining the extractable quantity. Second, the costs tied to reserve exploration and extraction play an important role in incentivizing the pursuit of discovering and producing natural gas. Lastly, the costs associated with the construction of substantial infrastructures act as a constraining factor, regulating the flow of natural gas to distribution centers. All these elements collectively significantly influence the overall supply dynamics of natural gas.

Furthermore, the security of the natural gas supply is a significant concern for many countries. The alterations in a country's natural gas imports and the corresponding decrease in exports can influence the security level of its supply system, as noted by Zhang et al. [2015]. Moreover, increasing the gas storage capacity can also enhance the security of supply by providing resistance to interruptions from other transport systems [Madžarević & Crnogorac, 2022].

Diversification of natural gas suppliers is another important aspect of enhancing supply security [Biresselioğlu et al., 2015]. A country can create a more resilient and adaptable natural gas supply network by engaging with multiple suppliers. In terms of future trends, the consumption of natural gas is expected to grow, driven by the interest of the European Union in diversifying gas sourcing and increasing energy supply security [EU, 2023]. Policies are being implemented to allocate resources along the natural gas supply chain and develop infrastructure to mitigate the risk of provision shortages. Lastly, pricing reforms have been instrumental in encouraging natural gas supply from domestic and overseas sources [Zhang & Paltsev, 2016].

Numerous studies have examined the determinants of natural gas demand in various countries and regions. Economic factors play a significant role in driving natural gas demand. Studies have found that income elasticity of demand, which measures the responsiveness of demand to changes in income, is a significant factor in natural gas consumption [Burke & Yang [2016], Dong et al. [2019]]. As income levels rise, there is an increased demand for energy, including natural gas, particularly in developing countries [Dong et al., 2019]. Additionally, price elasticity of demand, which measures the responsiveness of demand to changes in price, also influences natural gas consumption [Burke & Yang, 2016]. Higher natural gas prices can decrease demand, as consumers may seek alternative energy sources or reduce their consumption [Erdoğdu, 2010].

Environmental concerns and policies also impact natural gas demand. The shift towards cleaner energy sources and efforts to reduce greenhouse gas emissions have increased demand for natural gas due to its lower carbon emissions compared to other fossil fuels [Dilaver et al., 2014]. Government policies and regulations promoting the use of natural gas as a cleaner alternative to coal or oil can drive demand. For example, subsidies and incentives for natural gas consumption can stimulate demand [Ackah, 2014].

According to Dagher [2012], other factors influencing natural gas demand include technological advancements, energy efficiency improvements, and changes in energy use patterns. The availability and accessibility of natural gas infrastructure, such as pipelines and storage facilities, can also affect demand. Additionally, weather conditions, such as temperature, can impact the need for natural gas for heating purposes [Nick & Thoenes, 2014].

2.3 Pricing Mechanisms

Natural gas pricing is a complex process that is influenced by various factors. One crucial factor is the relationship between petroleum and natural gas prices. A study by Atil et al. [2014] found that oil prices have an asymmetric and nonlinear effect on gasoline and natural gas prices. However, the price transmission mechanism is different for both commodities. Furthermore, A. Brown & Yücel [2008] found that natural gas and petroleum products are substitutes, implying a continuum of prices between the two. This suggests that changes in petroleum product prices can also impact natural gas prices. According to a recent study by Rubaszek & Szafranek [2022], however, there are indications of a future trend where European gas prices are expected to decouple from oil prices, becoming solely influenced by market forces unique to the gas market.

In addition to the influence of crude oil prices, market reforms also play a role in determining natural gas prices. Wang et al. [2020] discuss how China, one of the world's largest natural gas importers, is reforming its pricing mechanisms to align domestic prices with international gas prices. This suggests that market reforms can help establish a more market-oriented pricing mechanism. The effectiveness of natural gas pricing reforms is a topic of discussion. Boqiang & Zhensheng [2020] argue that while pricing reforms in China have achieved desired results in the non-residential sector, the current mechanism does not accurately reflect supply and demand dynamics.

The impact of hedging transactions, taxation, and accounting practices on natural gas pricing is discussed by Çelenk & Bozlak [2022]. These factors can help control costs within a specific band interval or fix prices at a single level, reducing the effect of price fluctuations on natural gas costs. Furthermore, the behavior of natural gas exporters can affect natural gas prices. Heidari [2020] found that increased competition by gas exporters leads to downward pressure on natural gas prices.

Overall, natural gas pricing is influenced by various factors, including crude oil prices, market reforms, government policies, hedging transactions, taxation, and supply and demand dynamics. Establishing an effective pricing mechanism that reflects market conditions and considers environmental and economic factors is crucial for ensuring efficient and fair natural gas pricing.

2.4 Trading and Exchanges

Natural gas trading involves the buying and selling of natural gas in various forms, such as pipeline gas or LNG. Several factors influence the dynamics of natural gas trading. As highlighted by Shi [2016], a critical aspect of natural gas trading is the establishment of trading hubs, with some of the notable ones being Title Transfer Facility (TTF) in Europe, Henry Hub in the United States, and National Balancing Point(NBP) in the United Kingdom. These hubs serve as focal points for trading activities, providing benchmarks for pricing and aiding in the overall efficiency of natural gas allocation.

Traders frequently reference the prices established at these hubs, influencing the broader global dynamics of natural gas trading. According to Cai & Wu [2021], creating trading hubs in Asia and Europe is crucial for achieving more efficient gas allocation and ensuring that gas pricing reflects the fundamentals of the gas markets. Developing these hubs allows for competitive pricing and facilitates natural gas trading. The transition from oil indexation to competitive pricing is a crucial aspect of natural gas trading, as discussed by Shi et al. [2019]. The move towards competitive pricing involves the determination of potential gas trading hubs supporting competitive prices.

Risk management is another essential aspect of natural gas trading. Morkunas et al. [2019] emphasize that natural gas trading companies must assess and manage business risks. In a liberalized market, these companies may need to promptly purchase the deficient quantity of natural gas in the spot market to fulfill their contract responsibilities. Long-term contracts and destination restrictions also play a role in risk management. Shi & H. [2016] note that natural gas, particularly LNG, is often traded under rigid long-term contracts with destination restrictions. However, Xia et al. [2022] highlights the evolving nature of the natural gas market, with the rapid development of short-term and spot trade of LNG leading to a shift from regionalization to globalization.

In summary, natural gas trading involves the establishment of trading hubs, transitioning to competitive pricing, risk management, long-term contracts, destination restrictions, and price convergence. These factors shape the dynamics of natural gas trading and contribute to the efficient buying and selling of natural gas in the global market.

2.5 Current Natural Gas Market

This section derives insights from the reports published by IGU [2023] and Energy Institute [2023].

As outlined in Section 2.1, natural gas is traded either in its gaseous form through pipeline networks or in its liquefied state as LNG. Figure 2.1 illustrates the distribution of gas exports between LNG and pipeline transportation. The figure highlights a decreasing trend in total gas exports from 2021 to 2022, accompanied by a shift toward increased LNG exports. The factors contributing to this trend will be addressed later in this section. Furthermore, due to the distinct infrastructure required for exporting gas in various states, LNG's principal importers and exporters will vary from those involved in the import and export of pipeline gas [IGU, 2023].

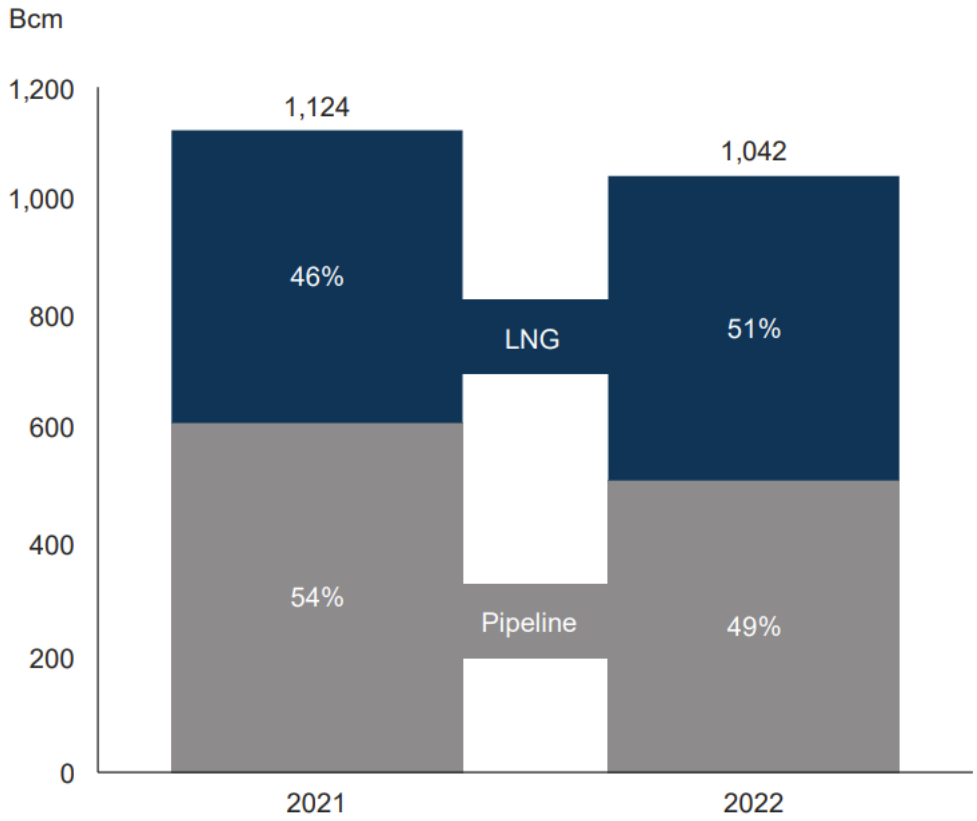


Figure 2.1: Percent gas volumes exported as LNG versus Pipelines.

Source: IGU [2023]

Figure 2.2 displays the gas flow for LNG trading in 2022. According to IGU [2023], the biggest exporters of LNG were Australia, the United States, and Qatar. Japan and China ranked as the largest importers, with imports of 73.6 million tonnes (MT) and 63.7 MT, respectively. China, however, experienced a notable drop of nearly 20 percent compared to 2021. In 2022, Europe witnessed a significant increase in LNG demand, resulting in imports of 126.6 MT. Furthermore, as of April 2023, the global LNG fleet comprises 668 active vessels with 312 new vessels under construction. Following the liquefaction and transportation of gas, the subsequent critical factor is receiving and regasifying the gas. The capacity of regasification terminals within each region will determine the limitations on how much LNG a country can import. Today, Japan, South Korea, and China possess the most extensive regasification capacities, with 217, 141, and 100 million metric tons per annum (MTPA). Within Europe, Spain, The United Kingdom, France, and the Netherlands possess the highest regasification capacities.

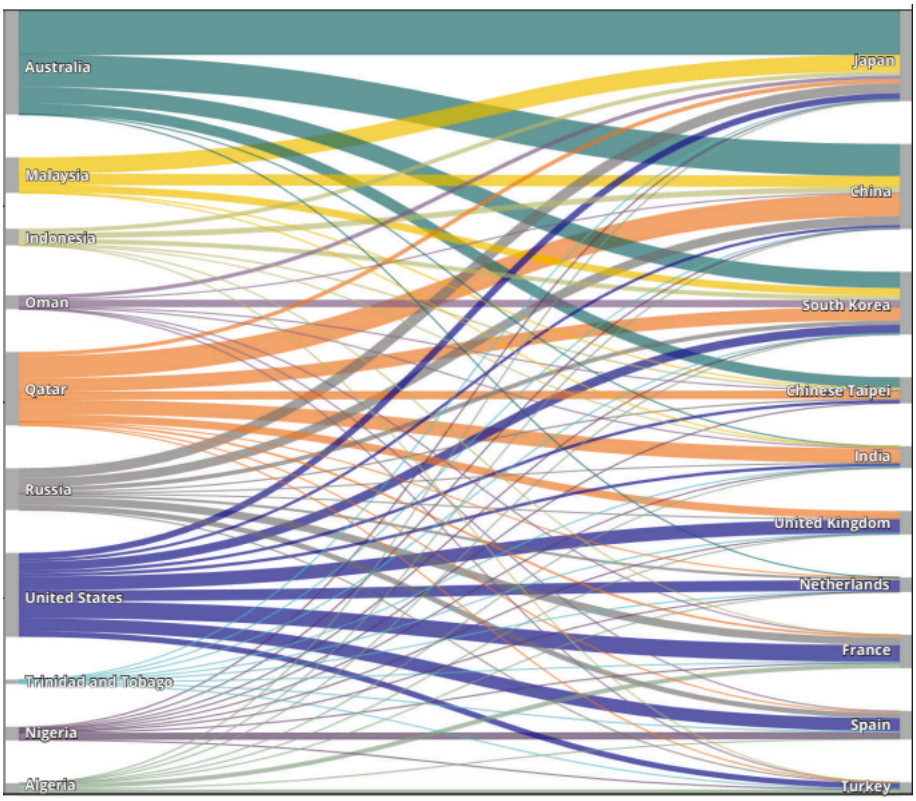


Figure 2.2: The flow of LNG trade in the market from exporters on the left to importers on the right.

Source: IGU [2023]

Per Energy Institute [2023], gas traded via pipelines decreased by approximately 15 percent in 2022. Figure 2.3 illustrates the flow of natural gas from significant exporters to major importers through pipeline networks. Despite a 29 percent decline in its global pipeline exports in 2022, Russia continues to be the largest exporter. Norway is the world's second-largest exporter of pipeline gas, accounting for a significant 25 percent of global exports. Europe is the biggest importer of natural gas by pipeline, with United States and China being the second and third biggest, respectively.

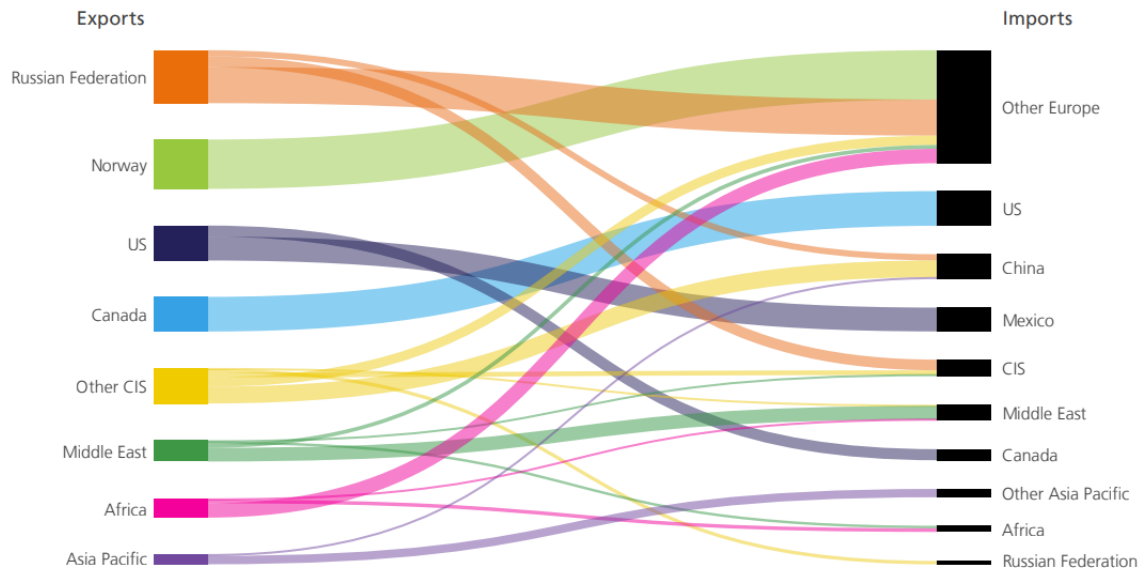


Figure 2.3: The flow of pipeline gas from exporters on the left to importers on the right.

Source: Energy Institute [2023]

Figure 2.4 illustrates the disparity in gas prices across different regions. As outlined by IGU [2023], it is evident that European gas prices exhibit higher volatility and, on average, higher prices compared to other regions. This trend is also observable in Asia and the Asia Pacific region. Regions with a greater reliance on LNG tend to exhibit heightened volatility. This is primarily because LNG import prices are subject to more bidding and are more susceptible to external factors like weather, fuel prices, and entry point capacity, in contrast to the relative static pipelines.

Per IGU [2023] the year 2022 marked one of the most turbulent periods in the history of the gas market, characterized by substantial supply and price shocks. The gas supply remains constrained in 2023, rendering the market highly sensitive to shifts in both supplier and demand changes. Following the Russian-Ukrainian conflict, Europe faced a significant reduction in its gas supply from Russian pipelines. The number of operational pipeline gas supply routes from Russia decreased from six to two, and this situation remains unchanged as of September 2023. This equates to a 34% reduction in the volume of gas imported to Europe. Consequently, Europe substantially increased its reliance on LNG to compensate for the gas shortage. This shift resulted in a 69% increase in LNG imports, positioning Europe as the largest importing market for liquefied natural gas. To address this heightened demand, the United States, being the primary supplier, augmented its LNG exports to Europe by 159% since 2021. Qatar, Russia, and Nigeria are the other key suppliers of LNG to Europe.

IGU [2023] asserts that while this adaptation has played a significant role in meeting the energy demand in Europe, it has also triggered an escalation in gas prices. In 2022, gas prices reached extreme heights and displayed extraordinary volatility, with the Western European and Northeast Asian markets reaching peaks of approximately 90 USD/MMBtu and 60 USD/MMBtu, respectively. This shift has had regrettable consequences regarding emissions, with China and India, for instance, intensifying their reliance on coal to address energy security concerns. Coal consumption for power generation in Europe rose by 1.3%, while in Asia, it increased by 2.6%, reflecting the gas-to-coal switching effect.

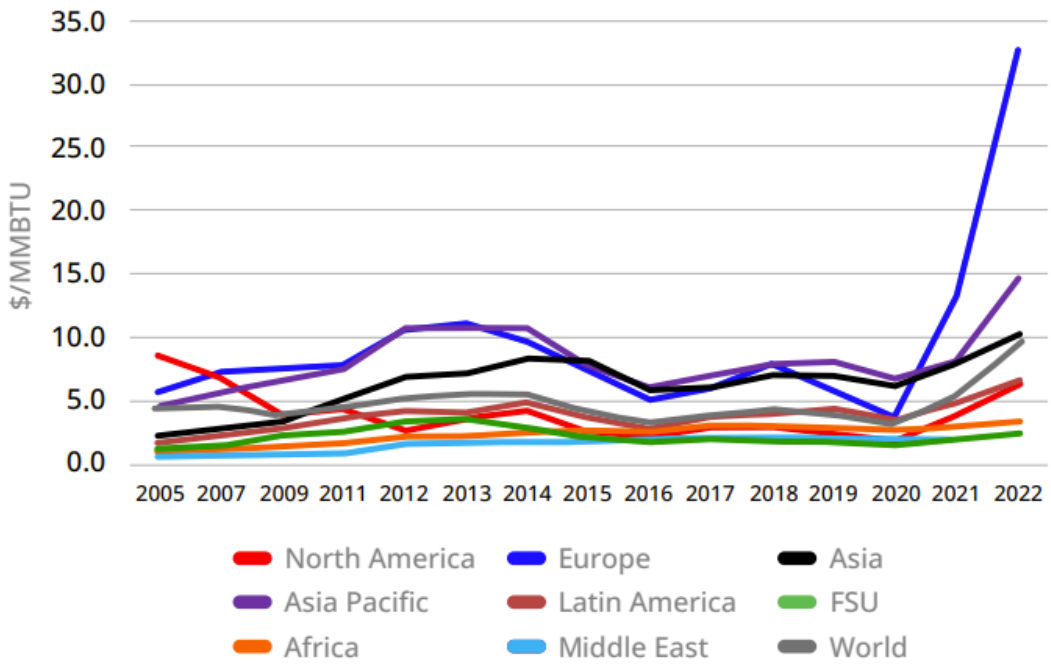


Figure 2.4: Gas prices by region.

Source: IGU [2023]

As per IGU [2023], throughout 2023, natural gas prices have primarily followed a downward trajectory. This descent can be attributed to reduced demand and a relatively mild start to the northern autumn, contributing to price declines. The return of Freeport LNG in the United States to production in February 2023, following an outage caused by a fire in June 2022, has also contributed to easing the pressure of gas prices in Europe and Asia by restoring a significant share of liquefaction capacity. Although prices are significantly lower than peak periods of 2022, prices remain volatile and are extremely sensitive to changes in market conditions. The Western European market continues to record average prices that are approximately three times higher and average volatility that is approximately five times higher compared to the levels observed before the COVID-19 pandemic. This heightened sensitivity primarily results from the exceptionally tight market balance, with no significant new supply additions expected in the next two years. The escalation of the conflict in the Middle East, coupled with the approaching cold season, is likely to further increase price volatility in the fourth quarter of 2023.

2.6 Natural Gas in Germany

Natural gas is Germany's second-most crucial primary energy source after petroleum. In 2022, its share of primary energy consumption amounted to 23.8% [Appun et al., 2023]. The Industry and Household sectors stand out as the primary gas consumers in Germany. Moreover, a substantial amount of natural gas is utilized for electricity generation [Kedzierski, 2023]. With the electricity generated from renewable sources varying considerably depending on weather conditions and season, natural gas-fired plants can play an important role in offsetting such fluctuations.

Over the years, Germany has established a considerable dependence on the import of natural gas, notably from Russia, where 55% of the country's imports originated in 2020 [Halser & Paraschiv, 2022]. This dependence proved problematic during the Russian invasion of Ukraine in 2022. In response, the government aims to reduce imports to 10% by 2024. To mitigate the risk of a gas shortage, Germany is seeking to increase the importation of LNG [Halser & Paraschiv, 2022]. Consequently, the natural gas infrastructure in Germany is undergoing expansion and development. The country has plans to approve up to 11 LNG terminals, both offshore and onshore, which will enable the importation of fossil natural gas until 2043 [Kemfert et al., 2022]. Currently, Germany fulfills a significant portion of its primary demand from Norway, constituting one-third of its total natural gas imports. Additionally, Germany addresses a portion of the supply gap through increased imports from Belgium and the Netherlands [Eckert & Steitz, 2023].

Starting in October 2011, Germany's natural gas market was partitioned into two distinct dual-quality market areas: NetConnect Germany (NCG) and GASPOOL (GPL). Nevertheless, as of June 1, 2021, GPL and NCG have been consolidated under the management of the newly established THE [Eckert, 2021]. Consequently, on October 1, 2021, Germany's unified nationwide gas market was established.

Various factors influence the pricing patterns in the German natural gas market. As per a study conducted in 2018, natural gas prices in Germany are predominantly linked to and indexed with oil prices [Zhang et al., 2018]. This indexing to oil prices is a common practice in this market. Moreover, an investigation into the determinants of natural gas prices in Germany conducted in 2014 revealed that market fundamentals, including supply and demand dynamics, contribute to determining prices [Nick & Thoenes, 2014].

	2021	2022	Change (%)
Production	50.4	47.2	-6.4
Imports	1673.3	1441.0	-13.9
Exports	768.9	536.0	-30.3
Net Imports	904.5	905.0	0.1
Storage Facility Balance	61.4	-86.0	-
Consumption	1016.3	866.2	-14.8

Table 2.1: Balance of German natural gas supplies in billions kWh

Source: Kedzierski [2023]

In 2022, Germany's natural gas consumption totaled 866.2 billion kWh, with domestic production meeting only 5.5% of the country's demand. Notably, the volume of this domestic production witnessed a decline of 6.3% compared to 2021, reaching 47.2 billion kWh. The nation experienced a notable decrease in both gas imports, which fell by 13.9%, and gas exports, which saw an even more substantial drop of 30.3%. Interestingly, in contrast to the previous year, 2022 saw a shift as more gas was injected into storage facilities than withdrawn from them, as reported by [Kedzierski, 2023].

Chapter 3

Forecasting Asset Prices: From Markowitz to Deep Learning

This chapter aims to guide the reader through the evolution of econometrics, spanning from theoretical frameworks to advanced machine learning algorithms. By presenting this historical evolution, we offer a deeper understanding of the motivation behind benchmark selections and provide motivation for exploring opportunities in new advancements.

For an extended period, financial analysts have strived to comprehend and forecast the future values of financial assets. In the early stages of econometric theory, the primary emphasis was on formulating theoretical frameworks to enhance the understanding of asset pricing. Markowitz [1952] introduced a pivotal mean-variance model, wherein investors actively sought to maximize their portfolio's expected return while considering a variance threshold. This framework operated on the assumption of investor efficiency. Building upon Markowitz's groundwork, Sharpe and Lintner developed the Capital Asset Pricing Model (CAPM) model in 1964 [Sharpe [1964], Lintner [1965]]. Within the CAPM, the appropriate expected return for an asset is determined by considering its systematic risk and its correlation with the market [Mossin, 1966]. Moreover, in addition to the assumptions made by Markowitz, Sharpe and Lintner assumed homogeneity of investor expectations and the ability of all investors to borrow or lend at a risk-free rate. Despite criticisms highlighting its limited accuracy in predicting asset returns in the market, the CAPM has garnered widespread popularity over the years. It persists as a fundamental tool for estimating the cost of capital [Elbannan, 2014].

Both the CAPM model and the work of Markowitz assumed the presence of an efficient market. Extending this premise, Fama [1991] introduced the groundbreaking efficient market hypothesis (EMH), a theory proposing that asset prices fully incorporate all available information. According to the EMH, achieving consistently above-average returns through the utilization of publicly available information is deemed implausible due to the swift and accurate adjustment of prices to new information. In essence, the EMH implies that it is not possible to consistently outperform the market by trading on publicly available information alone. There are different forms of the EMH: weak, semi-strong, and strong. The weak form suggests asset prices reflect all past market data, such as historical prices and trading volumes. The semi-strong form suggests asset prices reflect all publicly available information, including financial statements, news announcements, and analyst reports. The strong form suggests that asset prices reflect all information, including both public and private information. Critics of the EMH argue that markets are not always efficient

and that there are opportunities for investors to earn above-average returns. De Bondt & Thaler [1985] found evidence of stock market overreaction, suggesting that prices can deviate from their fundamental values in the short term. However, Fama [1991] argues that even if there are deviations from fundamental values in the short term, it is difficult to consistently profit from them due to transaction costs and the competition among investors.

Following the EMH, the belief held that all relevant information regarding financial assets was inherently embedded within their prices. Consequently, researchers concentrated their efforts on interpreting the statistical aspects of asset time series data. Coinciding with Fama's introduction of the EMH in the same year, 1970, Box and Jenkins offered their approach for modeling and utilizing the Autoregressive Integrated Moving Average (ARIMA) model [Box & Jenkins [2008], p. 88-103]. This approach marked a pivotal advancement in econometric theory. The ARIMA model combines autoregressive and moving average components to predict the future values of assets. The autoregressive component considers the correlation between an asset's value and lagged past values, while the moving average component considers the relationship between an asset value and past errors. By incorporating both components, the ARIMA model provides a flexible framework for modeling and forecasting time series data [Ho & Xie, 1998].

However, a limitation of the ARIMA model lies in its assumption of a constant variance within the series. Empirical evidence has demonstrated that this assumption does not hold for financial time series data [Morgan, 1976]. The emergence of the theory of heteroscedasticity in financial time series accelerated the development of more sophisticated models to capture the varying levels of volatility within these series accurately. In 1982, Robert F. Engle introduced the autoregressive conditional heteroskedasticity (ARCH) model, a non-linear framework for addressing heteroscedasticity in time series data [Engle, 1982]. This model acknowledges the presence of clustered volatility and incorporates the assumption that volatility is autocorrelated. It models the conditional variance in the error term using past squared error values.

Engle's efforts laid the foundation for developments by Bollerslev in 1986. He introduced the generalized autoregressive conditional heteroskedasticity (GARCH) model, which incorporates past lags of conditional variance into the model [Bollerslev, 1986]. The GARCH model and its variations have found extensive practical application [Wang et al., 2001]. Its ability to capture time-varying volatility has significantly impacted the field of econometrics, catalyzing the development of more sophisticated algorithms and methodologies.

Dixon et al. [2020, p. 4-6] writes that, in recent years, data has emerged as a precious asset, with an increasing abundance of new and more comprehensive datasets in the financial sector. Furthermore, "Alternative" data, encompassing information beyond traditional financial metrics, holds significant importance among market experts. For instance, social media data is considered a top source of alternative data, widely used by experts for market prediction. Many of these datasets are unstructured and are the foundation for intricate measurements. They often possess high dimensionality, incorporate non-numeric data, and exhibit complex non-linear dependencies, rendering them challenging for classical econometric models. Historical statistical models are susceptible to issues with high-dimensional data and often struggle to discern topological relationships [de Prado, 2019].

According to Dixon et al. [2020, p. 4-8], the convergence of new data challenges, coupled with the exponential growth in computing power and storage capacity, has ushered in a paradigm shift in econometrics towards big data techniques. Machine Learning techniques, empowered by increased computational capabilities, exhibit flexibility in recognizing more complex patterns within the data. Nevertheless, empirical evidence suggests that, especially for short-term forecasting, statistical methods can outperform sophisticated machine learning algorithms [Makridakis et al., 2018].

ANNs stand out as one of the most widely recognized categories of machine learning algorithms. The perceptron component, proposed by Rosenblatt [1958], marked a foundational step in their development. Subsequently, a cadre of researchers, including Widrow & Hoff, Ivakhnenko, Paul Werbos, Yann LeCun, and Geoffrey Hinton, among others, have made significant contributions to the continual evolution of neural networks [Prieto et al., 2016]. Dixon et al. [2020, p. 112-138] provides an in-depth description of Neural networks, which seek to replicate the intricate workings of the human brain through the use of perceptrons and synapse connections. Feed-forward Neural networks begin with the initial input of data into the input layer. This data traverses through the network, with each perceptron applying a transformation to its input through an activation function and weighing the inputs. In doing so, the network effectively operates as a mapping tool, converting inputs into outputs. The network receives feedback on prediction errors to enhance its accuracy, enabling the model to refine its weights and better align with the target values.

According to Prieto et al. [2016], neural networks received massive attention in the early 2000s and found real-world applications in various domains, including facial recognition, medical diagnosis, robotics, and financial time series forecasting. Following the popularity of artificial neural networks, numerous other machine learning techniques, including recurrent neural networks (RNN), tree-based architectures, and convolutional neural networks, have emerged. Notably, RNNs proposed by Elman [1990], and Long Short-Term Memory Networks (LSTM) proposed by Hochreiter & Schmidhuber [1997] have garnered significant attention, particularly in the context of time series forecasting. Apart from the standard feed-forward neural network input mechanism, they incorporate feedback connections, allowing predictions to account for prior time steps [Dixon et al. [2020], p. 239-255]. Fischer & Krauss [2018] showed that the LSTM outperformed memory-free classification methods when predicting directional movements of the S&P 500.

While recurrent architectures like LSTM and RNN have displayed potential in capturing sequential patterns, they may encounter challenges in capturing long-term dependencies [Yu et al., 2019]. In response to the limitations of recurrent architectures, a team of researchers, primarily from Google, introduced the Transformer model. As detailed in Vaswani et al. [2017], the Transformer model marked a significant departure from traditional transduction models by relying solely on attention mechanisms instead of RNN layers. The researchers contended that reducing the length of signal pathways for past input data makes detecting and learning long-term dependencies more feasible.

The transformer model was quickly adopted within the natural language processing community and is now considered state of the art for natural language generation models [Topal et al., 2021]. The architecture forms the fundamental basis of OpenAI's GPT-3 model. It is an exceptional natural language processing system known for comprehending the context of questions and producing high-quality responses [Ray, 2023]. Li et al. [2019] showed that the transformer model, with its ability to capture both long and short-term dependencies, is highly suitable for time series forecasting. The transformer model outperformed RNN-based methods when forecasting electricity production, wind and solar energy, and traffic data.

Building on the transformer model, Lim et al. [2021] proposed the TFT model in 2021. This innovative attention-based architecture integrates attention mechanisms designed to capture long-term dependencies, complemented by LSTM layers for localized processing. The TFT model extends the transformer architecture by considering static covariates and incorporating future known data, which makes it tailored for predicting future time series values. In the original paper, when compared to other methods, such as ARIMA and DeepAR, TFT consistently outperformed them, demonstrating performance enhancements ranging from 3% to 26%. The TFT model was quickly adopted and applied across various domains, yielding remarkable results [Wang et al. [2022], Nazir et al. [2023]]. Beyond its exceptional performance, the TFT model contributes to interpretability by providing insights into the relationship between input variables and target predictions.

The TFT model is one of the latest and most promising additions to the continuum of proposed models for time series forecasting. Its robust performance in comparison to alternative models makes it particularly intriguing for deployment in a financial context. Furthermore, the model addresses the challenging black box phenomenon, a common issue in many machine learning architectures. As explained by Guidotti et al. [2018], the lack of explainability introduces both ethical and practical concerns, and various approaches have been proposed to address this issue within established models. By incorporating interpretability into its architecture, the TFT model offers a fundamental solution to the challenge of explainability. The combination of high performance and interpretability makes the TFT model highly intriguing and could be a pivotal aspect of future advancements in the field.

Chapter 4

Data

This chapter presents the data employed by the models. Initially, we introduce each input feature and the corresponding motivation for their inclusion. Subsequently, we delve into selecting, collecting, and pre-processing the data to ensure its compatibility with our TFT model. Finally, we explore techniques employed for addressing missing values and the reshaping and standardization procedures applied to optimize the input data for our analysis. A complete list of all input features and their respective sources can be found in Appendix A.

4.1 Dataset

Significant research has been undertaken to identify effective input features for predicting future gas prices. In leveraging the comprehensive architecture of the TFT model, featuring distinct feature selection networks, we gather a diverse set of input features that have demonstrated the potential to enhance predictive accuracy. By supplying all data to the model, it can determine the inclusion or exclusion of features. Unfortunately, some desired data is not publicly accessible, rendering some input features less optimal.

Natural gas prices

The primary variable within the dataset is German gas prices, which serve as both the focus of our analysis and the target for our prediction. As detailed in Section 2.6, the consolidation of GPL and NCG into a single gas hub occurred in 2021. To create a unified time series for gas data, we gathered daily data from GPL and NCG from 01/01/2011 to 01/10/2021, and daily data for THE was collected from 1/10/2021 to 30/09/2023. Data from GPL and NCG were averaged to create a single composite gas price for Germany. This single price was merged with the THE prices to create one continuous German gas price throughout the period.

Figure 4.1 visualizes the continuous price data. The data displays a pattern consistent with the European market dynamics discussed in Section 2.5. Prices exhibited a sharp increase during the pandemic, particularly from 2021 onwards. Subsequently, prices surged following the Russian invasion of Ukraine in February 2022 before receding into an unstable equilibrium in the fourth quarter of 2022.

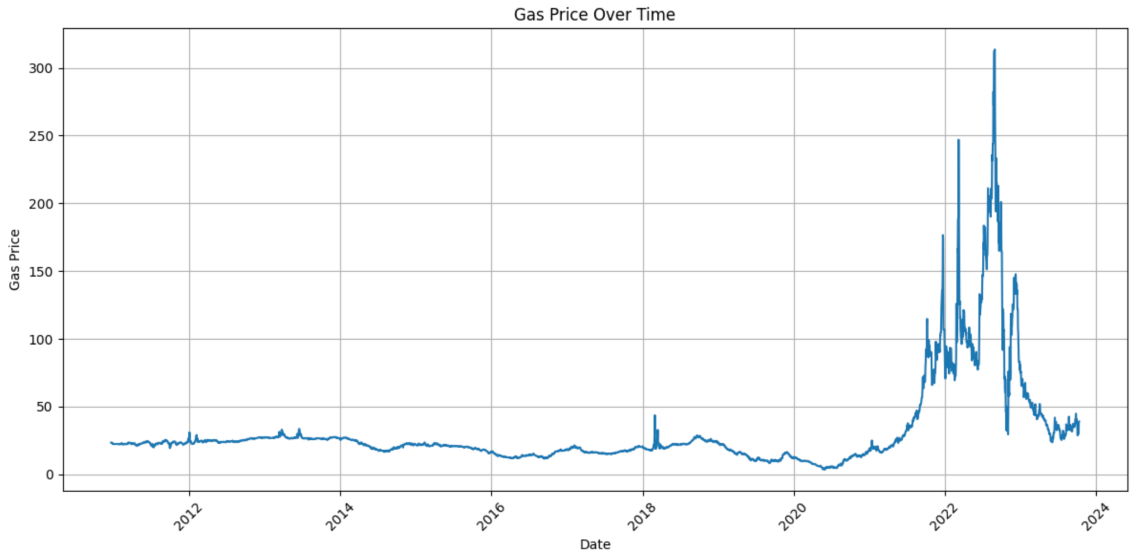


Figure 4.1: Germany natural gas price data. Average of GPL and NCG price data until 2021.

Source: Eikon

	German Gas	TTF	Henry Hub	NBP
German Gas	1.000000	0.993019	0.734612	0.924731
TTF	0.993019	1.000000	0.733773	0.932876
Henry Hub	0.734612	0.733773	1.000000	0.635841
NBP	0.924731	0.932876	0.635841	1.000000

Table 4.1: Price correlation between natural gas trading hubs

In addition to German gas spot data, we collect data from three other major gas hubs. As demonstrated in the study by Busse et al. [2012], the incorporation of TTF prices and NBP prices has significantly improved the predictive accuracy of German gas prices. Consequently, we include these features in addition to Henry hub prices. The Henry hub, situated in the USA, is one of the primary gas suppliers to Europe and holds a key role in global gas pricing. All prices are collected daily from 2011 to 2023.

Figure 4.2 illustrates the price trends for all trading hubs. TTF and German gas prices exhibit a strong correlation, closely tracking each other, whereas the Henry Hub consistently maintains a lower price throughout the period. NBP, on the other hand, displays a pattern similar to German gas but with greater volatility. Table 4.1 provides additional insight into the correlations between trading hub prices. TTF and German gas exhibit a robust correlation. NBP demonstrates high correlations with TTF and German gas, whereas the Henry Hub displays comparatively lower correlations.

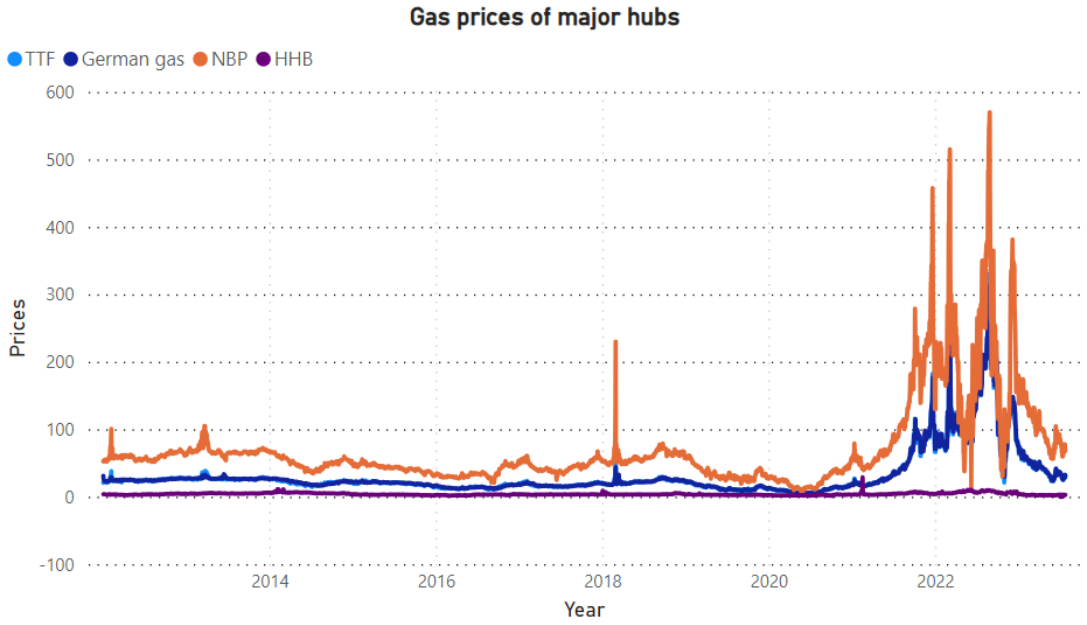


Figure 4.2: Prices per MWh for three major gas hubs in addition to German gas

Source: Eikon

Temperature data

In the research conducted by Busse et al. [2012], it was identified that temperature data played a significant role in predicting short-term price movements for German gas. As described in Section 2.6, gas plays a crucial role in heating in Germany, making the gas demand sensitive to temperatures. Incorporating temperature data in the model is, therefore, particularly interesting.

In their research, Busse et al. [2012] utilized historical weather forecasts ranging from one to five days ahead. Unfortunately, historical German weather forecasts are not freely accessible to the public; hence, we must depend on historical temperatures.

Historic temperatures are gathered for Germany's ten most populous cities. These temperature records are derived from daily aggregates, calculated by averaging hourly temperatures. Subsequently, this temperature data is consolidated to establish a daily average temperature for Germany. The constituent cities utilized in this aggregation process, weighted by their respective rounded population levels, are presented in Table 4.2.

City	Population
Berlin	3 800 000
Hamburg	1 900 000
Munchen	1 500 000
Cologne	1 100 000
Frankfurt	760 000
Dusseldorf	653 000
Stuttgart	610 000
Leipzig	624 000
Dortmund	610 000
Essen	593 000

Table 4.2: Ten largest cities in Germany

Source: Wartenburg [2023]

Figure 4.3 displays the average temperature data. The data indicates a seasonal pattern, as we would expect with cold winters and warm summers. January is the coldest month with an average temperature of 2.99°C, and July is the warmest month with an average temperature of 20.01°C. Analyzing across the years, 2012, 2013, and 2021 emerge as the coldest in terms of average temperatures, whereas 2020, 2022, and 2023 register as the warmest. Notably, 2022 exhibits an average temperature that surpasses that of 2013 by 3°C.

In addition to the weighted temperature index, we include an additional temperature index, simply the average daily temperature calculated from the cities in Table 4.2.

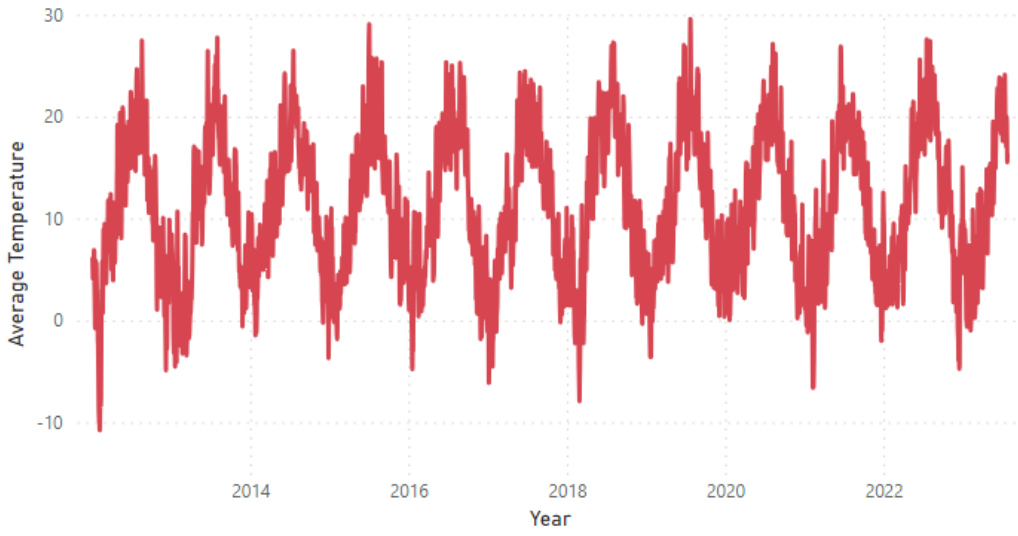


Figure 4.3: Average temperature of the ten largest cities in Germany

Source: Deutscher Wetterdienst

Other commodity data

In addition to identifying a notable impact of temperature shocks on German natural gas pricing, Nick & Thoenes [2014] emphasized the significance of long-term cross-commodity effects between German natural gas and oil and coal prices. Using a structural vector autoregressive model, they concluded that more than 60 percent of the variations in German gas prices could be explained by changes in oil and coal prices, making these features highly interesting in a machine-learning model. To capture these effects, we have collected Brent crude oil spot prices and one month ahead Rotterdam coal future prices from 2011 to 2023. Ideally, to capture the effects of changes in coal prices, we wanted to utilize data from the API 2 index, as undertaken by Nick & Thoenes [2014]. Unfortunately, this specific dataset was not freely available to the public.

German storage data

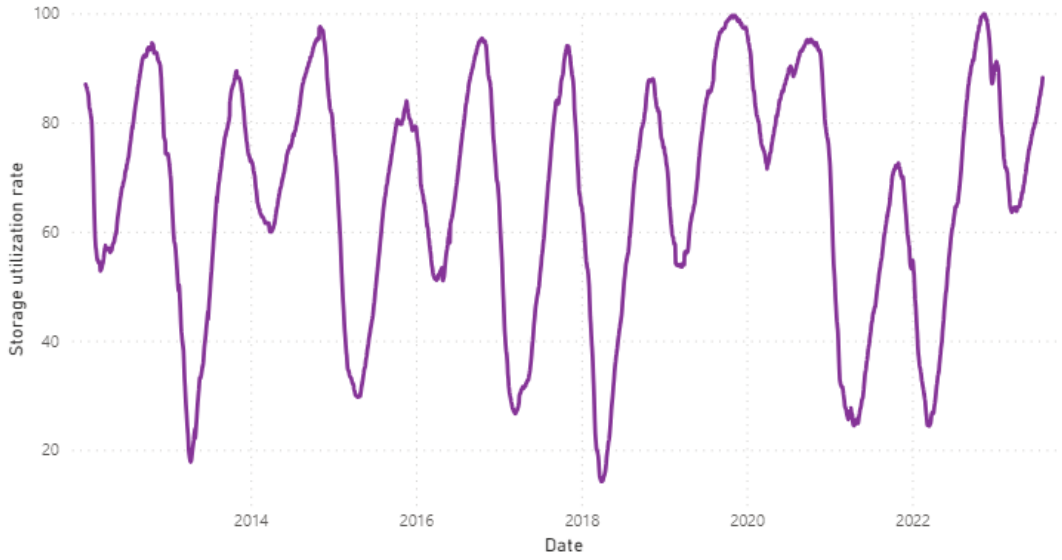


Figure 4.4: German storage utilization rate over time.

Source: Eikon

In a competitive market characterized by price fluctuations and highly seasonal demand patterns, utilizing storage facilities is crucial in balancing the dynamics between supply and demand. In addition to viewing storage activity as an indicator of market dynamics, Mu [2007] emphasizes that open storage information can incentivize investors to make decisions based on such insights, calling it the "storage announcement effect". The importance of this balancing activity has led many researchers to include storage activity as an essential feature in their models when predicting gas prices. Su et al. [2019] emphasizes using storage activity as a significant feature when predicting Henry Hub prices.

We gathered information on German natural gas storage to account for this data. Figure 4.4 displays the utilization rate of German natural gas storage. The storage reached its lowest point of 14.2% in March 2018. Notably, the year 2021 recorded the lowest average utilization rate of 47.95%. During a typical year, the storage is commonly utilized in the cold season and replenished during the summer, reaching its peak in October. We also incorporate daily storage activity through a net injection feature, calculated by subtracting daily GW/h natural gas withdrawals from GW/h natural gas injections to storage. The net injection will be negative when storage is emptied and positive when storage is refilled.

German electricity data

As detailed in Section 2.6, natural gas plays a crucial role in Germany's electricity generation. One can capitalize on these prices by purchasing natural gas and converting it into electricity during high electricity prices. Moradi et al. [2022] explores the arbitrage opportunity associated with this dynamic. To capture these effects, we therefore incorporate the electricity price. Figure 4.5 presents the German day-ahead electricity price. We observe a similar behavior to that of German gas prices, with extreme fluctuations from mid-2021 to 2023.

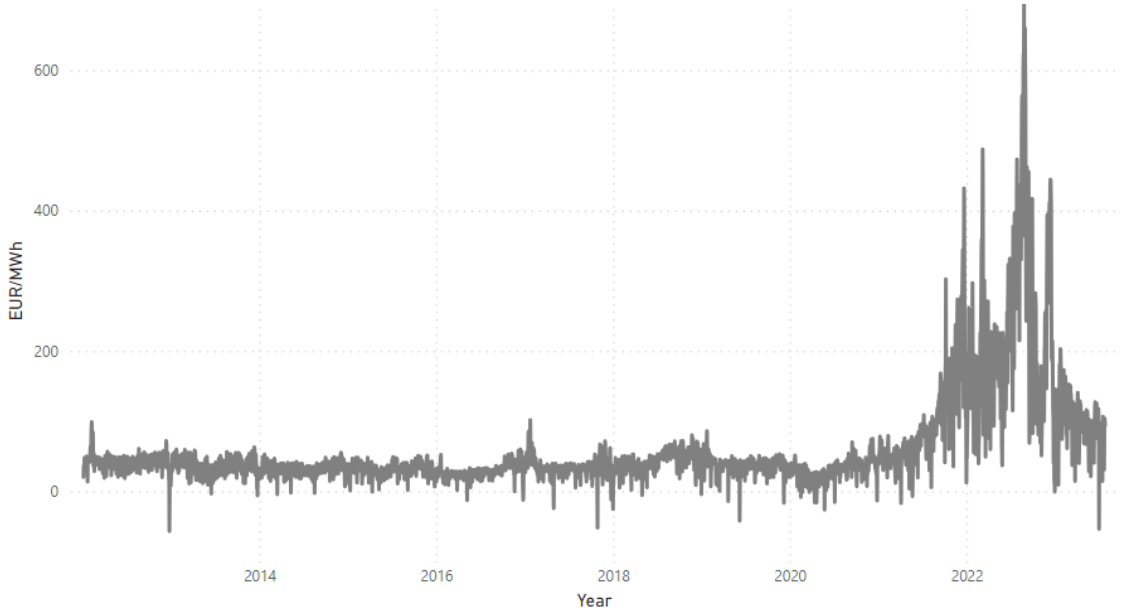


Figure 4.5: German day-ahead electricity prices from 2012 to 2023

Source: Energy-Charts

In addition to electricity prices, we incorporate electricity generation from wind and solar installations for the four electricity producers: 50Hertz, Amprion, TenneT TSO, and TransnetBW. Hulshof et al. [2016] investigated whether day-ahead predictions for German wind energy would affect TTF prices and found a positive significant relationship between the variables. In their research, they concluded that this relationship stems from bottlenecks created in the electricity grid when high levels of renewables are produced. These unscheduled flows are prioritized, leading to less cross-border capacity used for natural gas trade. Unfortunately, day-ahead predictions are not accessible for this data; however, we include realized daily values.

Google data

Tang et al. [2019] analyzes whether internet search data and news sentiment can significantly enhance forecasting abilities. Their study used an ANN to predict future NYMEX futures prices. Internet search data was collected from Google Trends for the word "Natural gas", and news sentiment from Yahoo Finance for the keyword natural gas. They demonstrated that incorporating this data into the model's input improved accuracy regarding Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), with Google data emerging as the most promising addition.

Considering the research conducted by Tang et al. [2019], we find it highly interesting to integrate internet searches into our model. The inclusion of this feature aims to capture certain political aspects related to natural gas, a dimension highlighted as necessary by Busse et al. [2012]. However, they faced challenges identifying a suitable data measure for this aspect.

To gather search data, we leverage Google Trends, a platform provided by Google that provides public information on search volumes on their search engine. The search volumes are normalized from 0 to 100, where 100 represents the highest search volumes within a specific period. The normalization process poses a challenge in data collection as all values are scaled relative to the highest value within the historical period from which the data is extracted. When examining historical data spanning over five years or more, only monthly data is accessible. Historical weekly data is available in five-year batches. To maximize data frequency while ensuring historical accuracy, we sample and normalize weekly data in five-year intervals and scale it based on monthly data. The normalization is based on the idea from Stejskalová [2023], who investigated whether Google searches can be utilized to predict stock returns within the automobile industry. Equation 4.1 normalizes the search volume indexed(SVI), capturing weekly fluctuations around the mean within each 5-year batch. The normalized weekly abnormal search volume indexes (ASVI) are multiplied with the monthly Google data, shown in equation 4.2 to capture correct scaling across the whole period.

$$ASVI_{Weekly} = \log(SVI_t) - \log(\overline{SVI_t}) \tag{4.1}$$

$$GoogleData_{Weekly} = ASVI_{Weekly} * GoogleData_{Monthly} \tag{4.2}$$

In line with the findings of Tang et al. [2019], we collect data for the search word "Natural gas" in Germany. The data is collected for the German area and consists of news searches. Furthermore, we also include data for the search word "War". Zaid & Khan [2023] concluded that the Russian-Ukrainian war of 2022 had significantly affected energy commodity prices. Moreover, Filis et al. [2011] notes that oil prices increased significantly during the second war in Iraq. These results suggest that integrating a signal of geopolitical tension could offer valuable information to the model.

Stock data

In the study by Kouchaksaraei et al. [2016], the impact of natural gas prices on the stock markets of Russia, Norway, and Qatar is investigated. The research utilizes monthly data from January 2005 to November 2013, encompassing natural gas prices and the stock exchange indices of the countries above. The findings reveal a two-way causality relationship between natural gas prices and the stock exchanges of both Russia and Norway.

In light of this research, our approach incorporates the stock prices of major energy companies from nations that export natural gas to Germany. Specifically, the stock price time series of Equinor, Gazprom, and ExxonMobil are included alongside the stock price of Eon, a prominent German energy conglomerate.

Additional data

In addition to the mentioned data, we also incorporate the EUR/USD exchange rate, CO_2 emission certificate spot, and shipping prices.

As detailed in Section 2.5, Europe has recently increased its reliance on LNG. Consequently, we aim to investigate whether including a shipping price index will add further value to the model. We integrate the Baltic Dry Index (BDI) into our analysis to explore this. As highlighted by Lin et al. [2019], the BDI serves as a short-term indicator for commodities, particularly during financial crises.

The exchange rate was identified as one of the five most crucial features by Busse et al. [2012]. This feature can act as an indicator for global economies and influence import prices from the United States, which, as highlighted in section 2.5, is one of Europe's major natural gas suppliers. Therefore, incorporating the exchange rate is highly interesting.

The day-ahead prices of CO_2 emission certificates are included as a feature and used in equation 4.3. As highlighted by Bai & Okullo [2023], there is a strong connection between CO_2 emission certificates and German gas prices. It is also suggested that the carbon allowance price can be one of the main drivers of coal-to-gas switching, making this variable of high interest.

4.2 Feature Extraction

In addition to collecting data, we extract new features based on the data collected in Section 4.1.

Spreads

In their study, Busse et al. [2012] explores spark spreads as input features, which experts highlighted as necessary. These spark spreads give the model an indication of the profitability of generating electricity from gas. We compute the feature CleanSparkSpread using Equation 4.3. This is a simplified version of the equation presented by Elias et al. [2016] where the cost of operating maintenance is not included. P_e represents the electricity price at time t, P_g is the German natural gas price converted to MMBtu (metric million British thermal units), HR is the heat ratio, P_c is the spot price of CO_2 emission certificates, and EF is ton CO_2 emitted per MWh produced by natural gas. As described in Section 4.1, the gas data is represented in EUR/MwH. To get an accurate estimate of the clean spark spread, we convert the price into EUR/MMBtu using Equation 4.4. According to S&P [2021], the conversion factor is 2.933. Efficiency rates and emission factor is used in line with Keles & Hasan [2020], where we use an average efficiency rate of 0.5 and an emission factor of 0.2.

A negative value of the clean spark spread might indicate an arbitrage opportunity where companies with generation capacities can buy excess natural gas, generate electricity, and sell electricity at a higher price.

$$CleanSparkSpread_t = \tag{4.3}$$
$$P_{et(EUR/MwH)} - P_{gt(EUR/MMBtu)} \cdot HR_{(MMBtu/MwH)} - P_{ct(EUR/tCO_2)} \cdot EF_{(tCO_2/MwH)}$$

$$P_{gt(EUR/MMBtu)} = P_{gt(EUR/MwH)} \cdot 2.933 \tag{4.4}$$

As outlined in Section 4.1, a significant correlation exists between German gas prices and other trading hubs. To facilitate the model's analysis of this relationship, we generate Hub spreads using Equation 4.5. In this equation, we calculate the spread by subtracting the German gas price from the prices of other hubs. Figure 4.6 illustrates the TTF spread, revealing a mean-reverting process where short-term deviations from the mean are corrected. By incorporating this relationship, the machine learning model can more easily find correct patterns in the data.

$$HUBSpread = GasPrice_{hub} - GasPrice_{Germany} \tag{4.5}$$

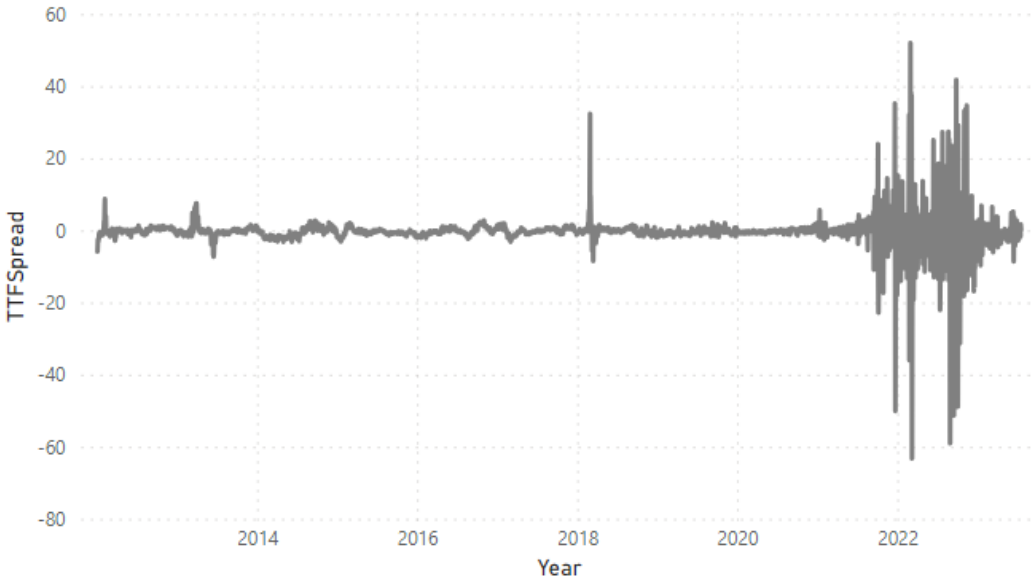


Figure 4.6: Spread between TTF prices and German gas prices

Source: Eikon

Date values

Following the approach outlined by Lim et al. [2021] in their proposal of the TFT model, we create new date features to capture seasonality within the data. Notably, when forecasting day-ahead prices for the TTF hub, Berrisch & Florian [2022] identified a significant Monday effect, indicating heightened price volatility on this day. Therefore, incorporating Day-of-the-week features into the model becomes particularly interesting. Additionally, in alignment with Lim et al. [2021], we introduce features for the month and day of the month.

Indicators

To enhance the model's ability to identify patterns in the data, we create features that signify specific trends. An essential trend for the model to capture is volatility. Volatility tends to occur in clusters, indicating that an observed increase in volatility can accelerate further price movements. This effect is especially interesting as Goor & Scholtens [2014] found an inverse leverage effect when analyzing UK gas prices, meaning an increase in volatility might lead to positive returns. We use R^2 as a volatility indicator. The R^2 is derived by squaring the outcomes of Equation 4.7.

In addition to incorporating the R^2 measure for German gas prices, we also introduce an R^2 feature for the oil price. As demonstrated by Lin & Li [2015], there is evidence of volatility spillover between Brent crude oil and European natural gas prices. This implies that an increase in oil volatility could result in heightened price fluctuations for natural gas.

As described in Section 4.1, temperature can have a significant predictive power in machine learning models. To capture unusual events in temperatures, we introduce a relative temperature indicator. This is computed using Equation 4.6, where T_t represents the temperature at time t, and $AvGT_t$ denotes the historical average for that specific day. Including this feature enables the machine learning model to detect whether the winter is exceptionally mild, potentially influencing seasonal patterns.

$$RelativeTemperature_t = T_t - AvGT_t \tag{4.6}$$

4.3 Data Preprocessing

Real-world data often contains noise in the form of missing values or unsuitable formats. Preprocessing steps are necessary to make the raw data suitable and optimal for the models.

4.3.1 Missing values

In our time series datasets, the presence of missing values is notable. This challenge is particularly pronounced because of the weekly frequency of data, such as Google data, alongside a significant portion of the data with a daily frequency. The machine learning models need all data to be on the same frequency to make predictions. To get daily values for all features, we forward fill values within each week. Forward filling is used so that no information is leaked into the future.

The issue of missing values is also particularly prevalent when dealing with financial time series. Since trading hubs are commonly closed on Saturdays and Sundays, data for these days are often absent. To create a unified dataset with daily values throughout the date range, forward filling is employed, whereby the values for these days are populated based on the preceding Friday's values.

4.3.2 Model preperation

The study conducted by Dixit & Jain [2021] showed that the accuracy of machine learning models is negatively affected when the target variable displays non-stationarity. To address this, we employ Equation 4.7 on the Gas price variable, calculating the log returns to ensure stationarity. Utilizing the Augmented Dickey-Fuller test on the transformed data, as suggested by Dickey & Fuller [1979], we obtain a test statistic of -11.9. This value significantly exceeds the 1% threshold of -3.43, leading us to reject the null hypothesis that the Target series contains a unit root.

$$Target_t = \ln\left(\frac{GasPrice_t}{GasPrice_{t-1}}\right) \tag{4.7}$$

A challenge arises when applying a machine learning model to a real dataset due to the varied scales in different time series. This diversity in scale representation can introduce bias in machine learning models, favoring features with higher scales. To address this challenge, various scaling techniques are available to normalize data so that all data range

within the same scale. The study conducted by Thara et al. [2019] demonstrated that the standard scaler outperformed most other scaling techniques when used in deep neural networks. Consequently, we employ standard scaling for all continuous input data. The standard scaler normalizes each feature, ensuring a mean of zero and a standard deviation of one. Equation 4.8 illustrates the scaling process where z is the transformed value of the feature, x is the feature value at time t, μ is the average value of x within the scaling horizon, and σ is the standard deviation. Each series is scaled using information exclusively from the training sets to prevent information leakage through scaling.

$$z_t = \frac{x_t - \mu}{\sigma} \tag{4.8}$$

Chapter 5

Methodology

This chapter outlines the methodology employed in our thesis, aimed at producing accurate and interpretable multi-horizon forecasts for German natural gas prices. Initially, we introduce the architecture of Temporal Fusion Transformers (TFT), detailing its diverse configurations, layers, and gating mechanisms. Following this, we present the benchmark models utilized. Additionally, we explore the segmentation of the time series of German natural gas prices into distinct periods. Subsequently, we expound on the training procedure. Finally, we outline the evaluation criteria employed in the assessment process.

5.1 Temporal Fusion Transformer

In this section, we thoroughly elucidate the TFT architecture's distinct components, visually depicted in Figure 5.1. The following section is derived from the research paper authored by Lim et al. [2021].

5.1.1 Multi-horizon forecasting

Multi-horizon forecasting is a crucial problem within time-series machine learning. Its utility extends to predicting essential economic indicators, encompassing important metrics like industrial production, GDP, and inflation rates across multiple future periods, as highlighted in the work [Prince et al., 2022]. Furthermore, multi-horizon forecasting is applied in financial prediction, where it is employed to anticipate stock prices and market trends across multiple future periods [Hawkes & Date, 2007]. These predictions aid investors, traders, and financial institutions make investment decisions and mitigate risks. Multi-horizon forecasting presents a challenge due to increased computational complexity and heightened uncertainty with the expansion of the prediction horizon. To maximize the accuracy of these forecasts, it is vital to incorporate all available information. Furthermore, a multi-horizon model should be able to incorporate information that may be known or unknown in the future. Moreover, the model needs to account for static variables, which are time-invariant.

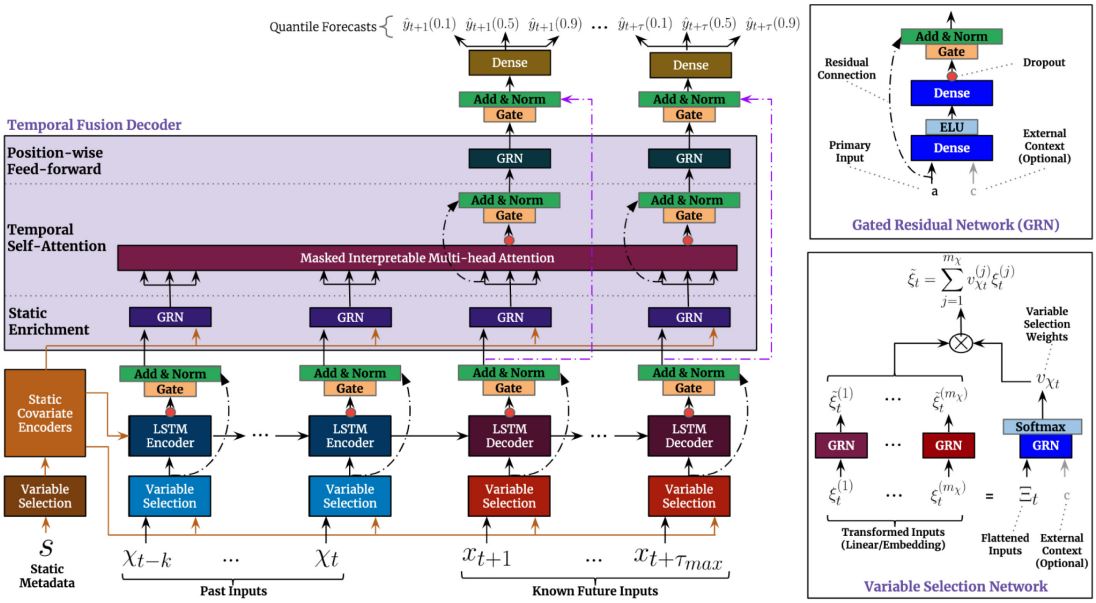


Figure 5.1: TFT processes three types of input data: static metadata, past time-varying inputs, and known future time-varying inputs. Variable Selection Network blocks are employed to emphasize the most significant features from these inputs. GRN blocks enhance information flow through skipping connections and gating layers. LSTMs analyze local patterns, and multi-head attention mechanisms integrate information across various timesteps.

Source: Lim et al. [2021]

Figure 5.1 visualizes the high-level architecture of TFT. We have I unique entities in the data. Each entity i has individual sets of covariates $\boldsymbol{s}_i \in \mathbb{R}^{m_s}$, as well as features $\chi_{i,t} \in \mathbb{R}^{m_\chi}$ and targets $y_{i,t} \in \mathbb{R}$ at each time-step $t \in [0, T_i]$. Time-dependent input features are subdivided into two categories $\chi_{i,t} = \left[z_{i,t}^T, x_{i,t}^T\right]^T$ - observed past inputs $z_{i,t} \in \mathbb{R}^{m_z}$ which are unknown beforehand, and known inputs $x_{i,t} \in \mathbb{R}_x^m$, which are known in advanced. The known features typically consist of date-related features that can, with certainty, be known in the future. TFT integrates historically observed inputs through a context window, incorporating all data from the start of the input window to the current day. The context representation, combined with known future inputs and static covariates, generates forecasts for each future day in the prediction horizon. This process is visualized in Figure 5.2.

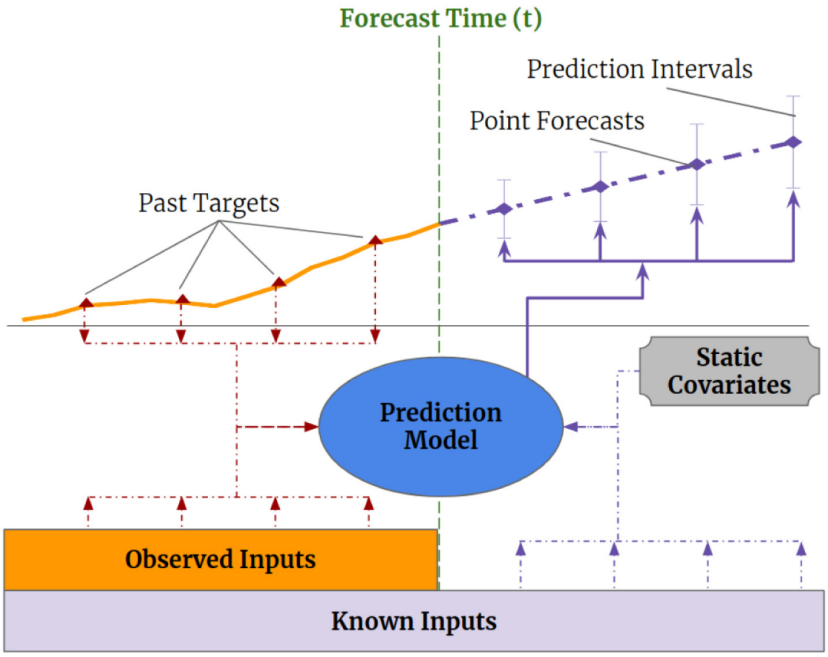


Figure 5.2: Multi-horizon forecasting with static covariates and various time-dependent inputs.

Source: Lim et al. [2021]

5.1.2 Gating Mechanisms

Some challenges in developing well-performing machine learning models include identifying the most relevant features and anticipating the level of complexity in the relationship between input features and targets. In specific scenarios, incorporating an extensive set of features introduces significant noise to the model. Additionally, during specific periods, unnecessary complexity in the model can lead to a high level of overfitting, resulting in inaccurate predictions.

To address this issue, the Gated Residual Network (GRN) is applied as a building block of TFT that allows the model to apply non-linear processing only when necessary. The GRN has a as its primary input and, depending on where the GRN is situated, uses static variables with a context vector c and yields.

$$GRN_{\omega}(a, c) = LayerNorm(a + GLU_{\omega}(\eta_1)), \tag{5.1}$$

where η_1 is defined as the following

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega}, \tag{5.2}$$

and η_2 is defined as

$$\eta_2 = ELU(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega}). \tag{5.3}$$

ELU is the Exponential Linear Unit activation as proposed by Clevert et al. [2015], η_1 and η_2 are intermediate layers. Furthermore, $LayerNorm$ is standard layer normalization, and ω is an index to denote weight sharing.

The TFT model also includes Gated Linear Units (GLU), offering the flexibility to bypass certain parts of the architecture when it is beneficial.

Letting γ be the input, the GLU takes the following form:

$$GLU_w(\gamma) = \sigma(W_{w,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega}), \tag{5.4}$$

where $\sigma(.)$ is the sigmoid activation function, W and b are weights and biases, $\odot$ is the element-wise Hadamard product. Both *ELU* and *GLU* aid the network in discerning simple input transformations from those demanding more intricate modeling by suppressing the non-linear calculations.

5.1.3 Variable Selection Networks

As previously stated, an extensive feature space can introduce substantial noise. The TFT model introduces Variable Selection Networks (VSN), enabling the model to learn the optimal selection of relevant features. Besides offering insights into the most significant variables, the TFT model aims to enhance performance by eliminating noisy features. Financial time series, such as gas prices, often contain significant noise, making filtering in the model highly interesting.

Consider $\xi_t^{(j)} \in \mathbb{R}^{d_{model}}$ as the transformed input for the jth variable at time t, and the vector $\Xi_t = \left[{\xi_t^{(1)}}^T, ..., {\xi_t^{(m_\chi)}}^T\right]^T$ the flattened vector of all past inputs at time t. Ξ_t and an external context vector c_s are fed through a filtering GRN unit and then a softmax function, producing a normalized vector $v_{\chi t}$.

Given the access to static information, the context vector c_s is omitted for static variables. Moreover, at each time step t, a supplementary layer of non-linear processing is applied by passing each ξ_t^j through its individual GRN. In this process, each variable is encoded using a neural network. Finally, each processed variable is weighted by its corresponding variable selection weight $v_{\chi t}^j$ and then combined, where $v_{\chi t}^j$ is the jth element of $v_{\chi t}$, making a single representation for each feature for all timesteps in the context window. Despite each variable sharing weights in the encoding process, the model can adjust to concentrate on specific historical time steps.

5.1.4 Static Covariate Encoders

TFT integrates information from static metadata by employing distinct GRN encoders, resulting in the generation of four distinct context vectors: c_s, c_e, c_c, and c_h. These context vectors are interconnected into different locations within the temporal fusion decoder. The variable c_s is employed in the variable selection process, c_c and c_h are utilized for local processing, and c_e is employed for enrichment. This process allows the model to increase the importance of one feature for certain instances while ignoring it for others.

5.1.5 Interpretable Multi-Head Attention

The TFT model utilizes a self-attention mechanism to learn long-range dependencies across different timesteps. Attention mechanisms scale values $V \in \mathbb{R}^{N \times d_v}$ according to the relationships between keys $K \in \mathbb{R}^{N \times d_{attn}}$ and queries $Q \in \mathbb{R}^{N \times d_{attn}}$. These terms are inspired by the retrieval process in information systems, where a query is used to search a database to retrieve the corresponding values. For example, considering the behavior of gas prices at time t, what are the typical values for weather? The process is employed after the feature selection encoding and is described as:

$$Attention(Q, K, V) = A(Q, K)V, \tag{5.5}$$

where $A()$ represents the normalization function. To enhance the learning capacity Vaswani et al. [2017] proposed multi-head attention, introducing multiple heads to project the input embeddings into different subspaces. The multiple attention head process is given by:

$$Multihead(Q, K, V) = [H_1, ..., H_{m_H}]W_H, \tag{5.6}$$

$$H_h = Attention(QW_Q^{(h)}, KW_K^{(h)}, VW_V^{(h)}), \tag{5.7}$$

where $W_Q^{(h)} \in \mathbb{R}^{d_{model} \times d_{attn}}$, $W_K^{(h)} \in \mathbb{R}^{d_{model} \times d_{attn}}$ and $W_V^{(h)} \in \mathbb{R}^{d_{model} \times d_{attn}}$ are head-specific weights for keys, queries and values, and matrix $W_H \in \mathbb{R}^{(m_H \dot{d}_V) \times d_{model}}$ condenses the concatenated matrices H_h into a single matrix.

Analyzing the attention weights of the model allows us to examine both the behavior of the model and the importance of the input features. However, when multiple heads are employed, interpretability becomes less clear. To enhance the interpretability of the attention, the TFT model modifies multi-head attention to share values across each head and utilize additive aggregation across all heads. The attention mechanism employed in the TFT model is then represented by:

$$\tilde{H} = \tilde{A}(Q, K)VW_H, \tag{5.8}$$

$$= \left\{ \frac{1}{m_H} \sum_{h=1}^{m_H} A\left(QW_Q^{(h)}, KW_K^{(h)}\right) \right\} VW_V, \tag{5.9}$$

$$= \frac{1}{m_H} \sum_{h=1}^{m_H} Attention\left(QW_Q^{(h)}, KW_K^{(h)}, VW_V\right), \tag{5.10}$$

where $W_V \in \mathbb{R}^{d_{model} \times d_V}$ are value weights shared across all heads and $W_H \in \mathbb{R}^{d_{attn} \times d_{model}}$ is used for final linear mapping.

5.1.6 Temporal Fusion Decoder

In the architecture, where the encoding process takes input features and maps them into the best possible representation, the decoder is utilized to learn the relationship between the encoded feature space, local values and the target values. In this process, a series of layers are employed.

Sequence-to-sequence layer

In time series data, identifying points of significance often involves considering their relationship with local surrounding values. Therefore, incorporating local context and understanding of sequential ordering can contribute to performance improvements. To accommodate various input types and the varying number of past and future inputs, the TFT architecture incorporates a sequence-to-sequence layer. This layer feeds $\tilde{\xi}_{t-k:t}$ into the encoder and $\tilde{\xi}_{t+1:t+\tau_{max}}$ into the decoder. The TFT model also incorporates a gated skip connection over this layer, defined as:

$$\tilde{\phi}_{t,n} = LayerNorm\left(\tilde{\xi}_{t+n} + GLU_{\tilde{\phi}}(\phi(t,n))\right), \tag{5.11}$$

where $n \in [-k, \tau_{max}]$ is a position index.

Static enrichment layer

TFT employs a static enrichment layer designed to augment temporal features with static metadata. Specifically, at a given position index n, the static enrichment process is defined as follows:

$$\theta_{t,n} = GRN_{\theta}(\tilde{\phi}(t,n), c_e), \tag{5.12}$$

where c_e represents a context vector obtained from a static covariate encoder. This layer introduces the relative feature importance, as previously discussed, where the significance of one feature can be observed in the context of a specific instance.

Temporal self-attention layer

In the same way as the encoder, the decoder employs interpretable multi-head attention for each forecasted time step given by:

$$B(t) = InterpretableMultiHead(\Theta(t), \Theta(t), \Theta(t))), \tag{5.13}$$

to yield $B(t) = \left[\beta(t,-k), ..., \beta(t,\tau_{max})\right]$. Decoder masking is then subsequently applied to ensure that no information leakage occurs. After the self-attention layer, an extra layer is introduced to aid in the training process:

$$\delta(t,n) = LayerNorm(\theta(t,n) + GLU_{\delta}(\beta(t,n))). \tag{5.14}$$

Position-wise feed-forward layer

The final part of the temporal fusion decoder is a position-wise feed-forward layer that applies additional non-linear processing to the outputs of the self-attention layer using GRNs:

$$\psi(t, n) = GRN_{\psi}(\delta(t, n)), \tag{5.15}$$

where the weights of GRN_{ψ} are shared across the entire layer. As seen in Figure 5.1, this layer is also directly connected to the LSTM layer through a gate that skips over the entire transformer block. In this manner, the model can learn to bypass unnecessary complexity.

Quantile outputs

In the domain of financial time series forecasting, it is not sufficient to merely predict the target variable; it is equally vital to assess the associated prediction uncertainty. TFT meets this requirement by generating probability quantiles for each prediction. These quantile forecasts result from a straightforward transformation of the output from the temporal fusion decoder:

$$\hat{y}_{q,t,\tau} = W_q \tilde{\psi}(t, \tau) + b_q, \tag{5.16}$$

where W_q and b_q are linear coefficients for the specified quantile q.

5.1.7 Interpretability Use Cases

The mentioned components above enable the analysis of individual elements to interpret the overall relationships the model has learned. The TFT architecture assists users in identifying globally important variables for the prediction problem and recognizing persistent temporal patterns. The use cases will be explored in this subsection.

Variable Importance

By analyzing the weights of the VSNs discussed in Section 5.1.3, we can distinguish important features from insignificant ones. During the training, the weights of the VSNs, denoted by $v_{\chi t}^{j}$ for each variable j, are aggregated.

The self-attention weights obtained from the interpretable multi-head layer unveil the attention importance of features through the context window. Analyzing these weights can provide valuable insights into the significance of input features. Combining Equation 5.8 and Equation 5.13, the self-attention layer contains the combined matrix $\tilde{A}$ of scores at each forecast time t. The outputs of the multi-head attention at each forecast horizon τ can then be characterized as a score-weighted sum of the preceding features at each position i:

$$\beta(t,\tau) = \sum_{i=-k}^{\tau_{max}} \alpha(t,n,\tau)\tilde{\theta}(t,i), \tag{5.17}$$

where $\alpha(t,i,\tau)$ is the (τ,i)-th element of $\tilde{A}$, and $\tilde{\theta}(t,i)$ is a sequence being processed.

Temporal Patterns

The attention weight patterns offer insights into the most crucial time steps that influence the decisions of the TFT model. In contrast to other models that rely on model-based specifications for seasonality and lag analysis, the TFT model possesses the capability to capture seasonality or lag effects through its own architecture.

5.2 Benchmark Models

To thoroughly evaluate the effectiveness of the TFT model, we perform a comprehensive comparative analysis with two prominent models: ARIMA and XGBoost. The subsequent subsections provide detailed explanations of both models. The hyperparameters applied to both models during training are presented in Appendix B.

5.2.1 eXtreme Gradient Boosting

Tree boosting stands out as an exceptionally potent and extensively employed machine learning technique. A noteworthy contribution to this field is XGBoost, a scalable machine learning system designed for tree boosting. This innovative approach was initially presented in the research paper authored by Chen & Guestrin [2016]. The methodology detailed in the following paragraphs is derived from the insights and techniques expounded in this work.

The fundamental concept of XGBoost is to minimize the following objectives:

$$L^t = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t), \tag{5.18}$$

where L^t is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_t$ and the target y_i.

The regulation term $\Omega(f)$ is expressed as the following:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2, \tag{5.19}$$

where T indicates the total number of trees, γ and λ are penalty coefficients, and w is a vector containing the score of each leaf. The second-order approximation is used to optimize the objective in general and we get:

$$L^t \backsimeq \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t), \tag{5.20}$$

where $g_i = \partial_{\hat{y}(t-1)} l(y_i, \hat{y}^{t-1})$ and $h_i = \partial^2_{\hat{y}(t-1)} l(y_i, \hat{y}^{t-1})$ are first and second order gradient statistics on the loss function. The constant terms are removed to obtain the following simplified objective at time t:

$$\tilde{L}^t = \sum_{i=1}^{n} (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2. \tag{5.21}$$

This Equation can be rewritten by expanding Ω as follows:

$$\tilde{L}^t = \sum_{i=1}^{n} (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t). \tag{5.22}$$

For a fixed structure $q(x)$, the optimal weight w_j^* of leaf j is given by:

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \tag{5.23}$$

and the optimal value is calculated as the following:

$$\tilde{L}^t(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \tag{5.24}$$

5.2.2 Autoregressive Integrated Moving Average

The following subsection is derived from Shumway & Stoffer [2017]. The ARIMA model integrates the differenced autoregressive and moving average models. Within this framework, the AR component signifies the regression of the time series against its historical data, while the MA component indicates that the forecast error is a linear combination of past errors. Lastly, the I component of ARIMA reveals that the original data values have been converted into differenced values of order 'd' to establish stationary data, a necessary condition for the ARIMA model.

$$y_t^{'} = I + a_1 y_{t-1}^{'} + a_2 y_{t-2}^{'} + ... + a_p y_{t-p}^{'} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + +\theta_q e_{t-q}, \quad (5.25)$$

where the predictors are the lagged p data points for the autoregressive part and the lagged q errors are for the moving average part. The prediction is the differences y_t in the d^{th} order.

5.3 Period Segmentation

As shown in Section 4.1, the behavior of German gas prices has undergone significant transformations over the years. The past three years have witnessed substantial shifts from relatively stable prices to pronounced fluctuations and heightened prices, eventually reaching a state of unstable equilibrium. Evaluating the model exclusively with recent gas prices will yield results specific to only one particular state of the gas prices. Given the uncertainty surrounding future gas price behavior, our goal is to establish a model that demonstrates robust performance across the entire dataset, encompassing all distinct periods of price behavior. A potential approach involves incrementally training the model for each day and generating corresponding predictions for the entire dataset. However, this is not feasible within our constrained timeframe due to the computational demands of the TFT model. To address this challenge, we propose dividing the dataset into segments according to the price behavior. The model is trained for each segment using data up to day $T - 60$, where T represents the final day within each segment. Subsequently, the model is then tested on the last 60 days. This approach enables us to assess the model's robustness across various segments while maximizing the utilization of training data.

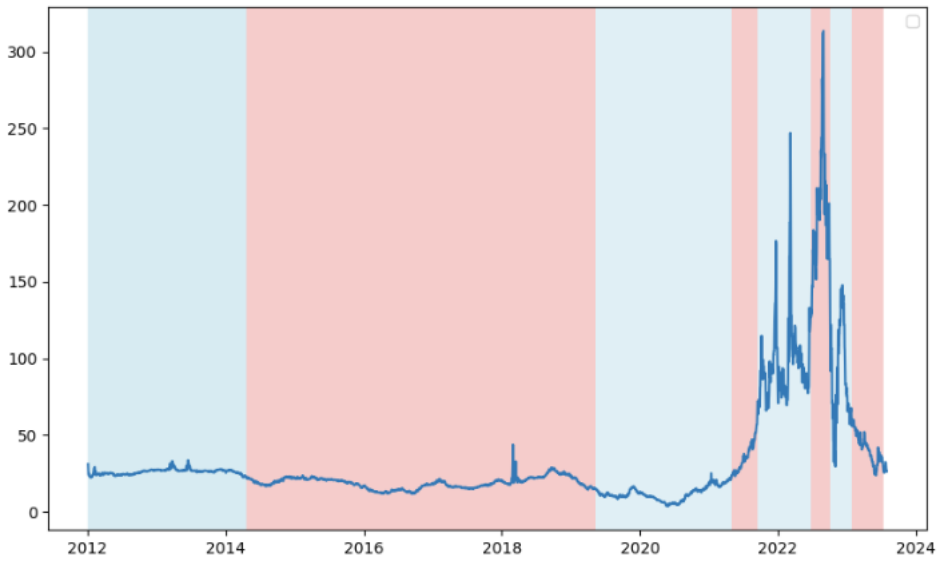


Figure 5.3: Change points over the period 2012-2023

We employ a change point detection algorithm to separate the time series of German gas prices into distinct segments. In their research, Van den Burg & Williams [2020] found that the binary segmentation algorithm proposed by Scott & Knott [1974] yielded the best results for univariate time series when default parameters were used. Utilizing the binary segmentation algorithm from the ruptures library by Truong et al. [2020], we identified eight distinct segments as depicted in Figure 5.3. The length of these segments varies, with the shortest segments observed in recent years.

The initial segment started in January 2012 and extended until April 2014, when a significant downturn in oil prices occurred, primarily attributed to supply-related factors. These factors encompassed the surge in U.S. oil production, diminishing geopolitical tensions, and changing policies within OPEC [Stocker et al., 2018]. The pivotal force behind the decline in natural gas prices was the substantial reduction, amounting to a 50% drop, in the cost per barrel of crude oil. However, multiple additional elements contributed to this trend. Economic downturns in Europe and Asia and the global adoption of more fuel-efficient vehicles collectively resulted in a decreased demand for gasoline [Isidore, 2014].

The second segment starts in April 2014 and extends until May 2019. Notably, 2018 witnessed record-high natural gas demand, propelled by increased reliance on natural gas-fired generation and the growth of LNG exports [Anderson, 2019]. Furthermore, in 2019, a sharp decrease in gas prices was observed [Shelor, 2020]. This decline was influenced by various factors, including a slowdown in China's economic growth to an estimated 6.1%, as well as reduced consumption in Eurasia and South America [IEA, 2019].

The third segment, starting in May 2019 and extending until April 2021, is marked by noteworthy warm weather patterns and a contraction in economic activity attributed to the COVID-19 pandemic [EIA, 2021]. This economic downturn led to 2020 natural gas prices at the benchmark Henry Hub plummeting to their lowest level in 25 years. Throughout 2020, prices steadily declined into the summer, influenced by a reduced LNG export demand and decreased heating demand due to a warm spring. However, the latter half of the year witnessed a resurgence in LNG demand, contributing to a strengthening of prices [Holland, 2021].

In the fourth segment, which starts in April 2021 and ends in September 2021, natural gas prices surged globally as a result of the economic recovery from the impacts of COVID in 2021. European prices, in particular, experienced a dramatic escalation when the major external supplier, Gazprom, initiated supply withholdings in Q4 2021. Additionally, 2021 witnessed a substantial rise in global demand for natural gas, while various geopolitical, environmental, and economic factors concurrently constrained the overall supply [Kotek et al., 2023].

Segment five, which starts in September 2021 and ends in June 2022, is distinguished by the Russian invasion of Ukraine. The invasion in February 2022 introduced substantial geopolitical and market uncertainties, particularly regarding the supply of Russian natural gas, culminating in a significant and abrupt surge in natural gas prices later that month. This escalation compounded pre-existing tensions within the natural gas market. After the invasion, reductions in Russian pipeline gas supply during 2022 triggered a rapid succession of price increases in European natural gas futures, ultimately reaching record levels in late August [ESMA, 2023].

Segment six, starting in June 2022 and ending in October 2022, is marked by a substantial decline in natural gas prices attributed to various factors. Initially, concerns about shortages prompted the European Council to set targets for member states to fill gas storage facilities to a minimum of 80 percent capacity before winter, a goal surpassed ahead of schedule. Concurrently, domestic gas resources typically increased in the months leading up to winter, with average reserves in October registering a year-on-year increase of approximately 15 percent [HSN, 2022]. Secondly, milder-than-expected weather across Europe reduced heating demand, allowing for the conservation of the continent's gas inventories [IEA, 2022]. Lastly, the manufacturing Purchasing Managers' Index (PMI) figures from major steel-consuming countries indicated a decline, signifying a deterioration in activity within those sectors. This decline was further accentuated by several domestic steelmakers' idling of melting capacity in response to weakened steel demand [HSN, 2022].

Segment seven, starting in October 2022 and ending in March 2023, is characterized by a sharp downturn. This could be attributed to the EU natural gas price cap implemented on December 3, 2022 [Welle, 2022].

Segment eight, ending in June 2023, is marked by a significant downturn in natural gas prices, driven by a combination of factors. The presence of relatively mild temperatures, record production levels, and inventories surpassing the usual averages collectively contributed to the reduction in natural gas prices [EIA, 2023].

	Start Date	End Date	Min Price	Max Price	Standard Deviation	Net Change
p_1	2014-02-11	2014-04-11	22.3	26.24	1.12	-2.83
p_2	2019-03-07	2019-05-05	14.4	17.9	0.77	-2.65
p_3	2021-02-24	2021-04-24	16.0	21.9	1.67	4.43
p_4	2021-07-14	2021-09-11	34.26	56.9	6.15	22.14
p_5	2022-04-30	2022-06-28	77.3	132.9	17.1	28.9
p_6	2022-08-03	2022-10-01	164.6	313.6	41.4	-37.7
p_7	2023-01-10	2023-03-10	43.6	70.6	6.5	-18.7
p_8	2023-05-02	2023-06-30	23.7	41.9	4.6	-2.4

Table 5.1: Characteristics of each test period

Table 5.1 presents the characteristics of each test period, which comprises the last 60 days within each period denoted as $p_1, p_2, \ldots, p_8$. Period p_1 and p_2 exhibit a relatively small standard deviation, both characterized by a downward trend in gas prices. Period p_3 shows a comparatively more significant standard deviation with a net change in prices of 4.43. Periods p_4 to p_7 witnessed sharp increases in volatility, with period p_6 being the most extreme, featuring a standard deviation of 41.1. The last period demonstrates similar behavior as described in section 2.5, in an unstable equilibrium. Although it shares a similar net change pattern with period p_1, the standard deviation is notably higher.

5.4 Training Procedure

As outlined in the preceding sections, a pivotal component of the TFT model involves temporal processing, acquiring both lengthy- and short-term temporal relationships from observed and known time-varying inputs. Consequently, we leverage distinct segments detailed in Section 5.3 to facilitate the training and testing of the data.

5.4.1 Train test split

As mentioned in Section 5.3, the model is trained using data from day T=0 up to $T - 60$ for each segment before being tested on the corresponding test periods shown in table 5.1. Predictions are generated for each day within the test period, projecting forward for the subsequent 14 days ahead. This process yields $60 \cdot 14$ predictions for each segment.

5.4.2 Loss Function

TFT is trained by jointly minimizing the quantile loss summed across all quantile outputs.

$$\mathcal{L}(\Omega, W) = \sum_{y_t \in \Omega} \sum_{q_t \in \ell} \sum_{\tau=1}^{\tau_{max}} \frac{QL(y_t, \hat{y}(q, t-\tau, \tau), q}{M_{\tau_{max}}} \tag{5.26}$$

$$QL(y, \hat{y}, q) = q(y - \hat{y})_+ + (1 - q)(\hat{y} - y)_+, \tag{5.27}$$

where Ω is the domain of the training data containing N samples. W represents the weight of TFT, and $\ell = [0.1, 0.25, 0.5, 0.75, 0.9]$ are the 10th, 25th, 50th, 75th and 90th percentile. In the quantile regression loss specified in Equation 5.27, the first term becomes prominent and positive when the predicted value, $\hat{y}$, is less than the actual value, y, resulting in a penalty for under-predictions. Conversely, the second term dominates when $\hat{y}$ exceeds y, penalizing over-predictions.

When q is set to 0.5, under-predictions and over-predictions are equally penalized, yielding the median. As the value of q increases, the penalty for under-predictions intensifies compared to over-predictions. For example, at $q = 0.9$, under-predictions are penalized by a factor of 0.9, while over-predictions are penalized by a factor of 0.1. Consequently, the model endeavors to mitigate under-predictions approximately nine times more aggressively than over-predictions, resulting in the 0.9 quantile.

5.4.3 Parameter Optimization

The precise hyperparameters governing the learning process are selected from a predefined search space outlined in Table 5.2. Inadequate combinations of hyperparameters yield suboptimal results by failing to minimize the loss function. Consequently, hyperparameter tuning is a crucial component in optimizing any machine learning model.

Hyperparameter	Search Grid
Encoder	30,60,100,365
Learning rate	0.01,0.001,0.0001
Hidden size	32,64,128,256
Attention head size	1,4,8,10,14
Dropout rate	0.1,0.2,0.3
LSTM layers	1,2,3,4

Table 5.2: Hyperparameter search space.

The sheer abundance of hyperparameter values renders the search space for identifying optimal model configurations overwhelmingly complex. Employing a grid search, which tests every conceivable combination, is impractical given computational constraints. Consequently, hyperparameter optimization is executed through 50 iterations of random search. Furthermore, due to constraints in computational resources, hyperparameter tuning is conducted once on the entire dataset, excluding the test periods.

Every TFT model undergoes training on a dedicated Nvidia RTX A2000 GPU using Cuda. Our optimal TFT models require slightly over 9 hours for the entire training process for all segments, with each epoch lasting approximately 1 minute. The specific parameter specification of our trained TFT model is detailed in Table 5.3 below.

	Parameter	Value
Library details	pytorch-forecasting	version 1.0.0
Network details	Encoder	365
	LSTM layers	2
	Dropout rate	0.2
	Hidden layer size	128
	Attention head size	8
	Loss	QuantileLoss
Training details	Epochs	80
	Minibatch size	32
	Predictions	14
	Learning rate	0.003
	Max gradient norm	0.9
	Optimizer	RANGER
Computational cost	Hardware	Nvidia RTX A2000
	Minutes per average epoch	2

Table 5.3: Information on library and optimal TFT configuration.

5.4.4 Feature sets

To identify the most practical combination of features for our analysis, we have created various feature sets, each comprising a blend of input variables. Due to computational constraints hindering an exhaustive search for the optimal feature set, we have undertaken a manual approach to craft diverse configurations. The table below presents the feature sets designed and explored in this study, each representing a unique combination of factors that we anticipate will contribute to the precision of our analysis.

Name	Features
Simple	German Gas Prices
Base	Weather data, Gas Prices and Exchange data
Feature set 1	Base + Google data
Feature set 2	Base + Storage data
Feature set 3	Base + Electricity data
Feature set 4	Base + Commodity data
Feature set 5	Base + Stock data
Full	All Features
Pooled	Simple feature set pooled by Hub

Table 5.4: Feature sets

The feature sets presented in Table 5.4 encapsulate distinct combinations of input features designed for our analysis. The "Simple" feature set includes solely the "Gas Price" feature and temporal features such as Day of the week and Month. When researching optimal parameters for predicting German gas prices Busse et al. [2012], identified five features for best performance using an ANN. These features include trading hub prices, EUR/USD exchange rate, and weather data. These features form our "Base" in table 5.4. To test whether the TFT model identifies additional patterns compared to an ANN or if the dependencies have changed through the years, we test additional features through feature sets 1 to 5. Each feature set adds data from specific domains. The "Full" set incorporates all available data as inputs and assesses whether the variable selection networks of the TFT model can effectively choose the most optimal features. A complete list of all features associated with each feature set is presented in Appendix C.

Inspired by Lim et al. [2021], we additionally evaluate a pooled model. In this configuration, price data from each natural gas hub is consolidated into a single feature set named "Pooled". Leveraging the static enrichment layer, the TFT can discern common patterns across hubs while isolating effects unique to individual hubs. By comparing the accuracy of the "Pooled" feature set with the "Simple" feature set, we can assess whether the model demonstrates improved performance through data pooling.

5.5 Evaluation Criteria

To comprehensively evaluate the model, we introduce the diverse accuracy measures employed to identify deviations from the actual results. Finally, we elaborate on applying the Diebold-Mariano test to assess the statistical significance of potential improvements.

5.5.1 Accuracy Measures

The performance of each feature set discussed in Section 5.4.4 is assessed based on three accuracy criteria. We employ Price MSE, Target MSE, and Quantile loss. MSE is the average squared difference between the values observed in a statistical study and the values predicted from a model and is calculated according to Equation 5.28

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{5.28}$$

where y_i is the ith observed value, $\hat{y}_i$ is the corresponding predicted value for y_i, and n is the number of observations. Given that the model's objectives involve forecasting future returns, we compute the Target MSE directly from the model's predictions and actual returns using Equation 5.28. The predictions employed in this computation correspond to the 50th percentile of the quantile predictions detailed in Section 5.27.

Price MSE, however, differs slightly from target MSE. We have the forecasted price, which is given by Equation 5.29

$$\hat{P}_t = \hat{P}_{t-1} \cdot e^{1+r_{0.5t}} \tag{5.29}$$

where $\hat{P}_t$ is the forecasted price at time t, and $r_{0.5t}$ is the 50th percentile forecasted return at time t. The process involves forecasting returns and multiplying them by the last forecasted price to obtain a new price. When t is 0, $\hat{P}_{t-1}$ corresponds to the last observed price in the training period. Thus, the Price MSE is given by Equation 5.30

$$PriceMSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - \hat{P}_t)^2 \tag{5.30}$$

where P_i is the actual price at time t, $\hat{P}_t$ is the forecasted price at time t, and n is the number of observations. As the price depends on each individual prediction, the price MSE allows us to assess the model's capability to predict a target series rather than focusing solely on pointwise predictions.

Additionally, as outlined in Section 5.4.2, the TFT model is trained by minimizing the Quantile loss. As a third criterion for evaluating the performance of distinct feature sets in the TFT model, we assess the quantile loss using Equation 5.27.

In the evaluation between TFT and the benchmark models, the lack of quantile forecasts in the latter makes the comparison using the quantile loss criterion inconclusive. Hence, we assess forecasts within each period through Target and Price MSE. Additionally, we introduce the "Correct Way" measure as an evaluation criterion. This binary measure takes the value of 1 when the deep learning model successfully forecasts the direction of the 14-day price movement and 0 otherwise. This evaluation criterion tests the model's ability to recognize price movement patterns.

5.5.2 Significance test

To test the significance of the predictions, we employ the Diebold-Mariano test, as proposed by Diebold & Mariano [1995]. The null hypothesis (H_0) posits that the difference in prediction errors between two series of forecasts is zero. However, it is noteworthy that Diebold [2015] argued that the Diebold-Mariano test was designed for comparing forecasts rather than models.

In our study, each period encompasses 60 prediction days, with each day generating a 14-day forecast, resulting in a total of 840 data points for each period. Despite the caution recommended in using the Diebold-Mariano test, given the substantial number of data points calculated for each period, it still provides a meaningful indication of significance.

Chapter 6

Results

We will present our findings in three stages. Initially, we will evaluate the performance of each discussed feature set outlined in Section 5.4.4. Subsequently, we will assess the performance of the TFT model for each period, considering both comparisons with other models and an analysis of the model dynamics. Finally, we will explore the TFT model's capability to recognize temporal patterns in the data.

6.1 Feature Selection

		p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
Simple	Q	11.86	21.16	29.13	30.72	71.41	83.53	55.49	67.16
Model	MSEP	0.45	0.78	**0.66**	11.89	277.18	2222.1	17.78	22.83
	MSET	0.00011	0.00034	0.00034	0.00081	0.0032	0.0051	0.0020	0.0030
Base	Q	10.93	18.6	**21.4**	31.7	**59.2**	77.0	46.4	58.6
Model	MSEP	0.5	0.64	0.87	17.47	254.5	2247.4	54.6	24.6
	MSET	0.00011	0.00034	**0.00029**	0.0009	0.00034	0.0054	0.0018	0.0034
Feature	Q	11.6	19.2	26.48	27.9	59.5	79.9	**44.7**	62.1
set 1	MSEP	0.81	0.76	3.25	12.96	304.76	3644.7	19.30	22.60
	MSET	0.00012	0.00033	0.0006	0.00082	0.0033	0.0056	0.0020	0.0029
Feature	Q	11.8	19.6	28.2	35.3	59.8	77.9	52.9	59.6
set 2	MSEP	0.698	0.72	1.33	17.92	278.4	1974.31	**14.59**	**18.2**
	MSET	0.00012	0.00035	0.00035	0.00088	0.0032	0.0051	0.0017	0.0039
Feature	Q	12.7	19.9	24.3	31.2	59.9	**76.5**	52.5	58.3
set 3	MSEP	0.62	0.82	0.94	17.0	257.59	1995.29	22.02	21.60
	MSET	0.00012	0.00035	0.00032	0.00087	**0.0031**	0.0052	**0.0015**	**0.0028**
Feature	Q	12.89	20.3	27.57	33.8	61.5	79.4	51.4	58.8
set 4	MSEP	0.73	0.68	1.52	18.58	**250.97**	2023.75	17.78	22.1
	MSET	0.00012	0.00033	0.00037	0.00089	**0.0031**	0.0051	0.0017	0.0029
Feature	Q	12.4	20.5	25.9	30.6	60.8	77.2	48.3	67.7
set 5	MSEP	0.62	0.68	1.11	15.5	261.8.5	**1889.89**	18.07	22.2
	MSET	0.00011	0.00033	0.00032	0.00085	0.0032	**0.0048**	0.0017	0.0029
Full	Q	**10.5**	**17.9**	28.0	**27.3**	61.3	80.5	57.6	68.8
Model	MSEP	**0.45**	**0.55**	1.10	**10.16**	299.80	2029.51	48.6	56.5
	MSET	**0.00010**	**0.00032**	0.00033	**0.00079**	0.0033	0.0050	0.0033	0.0046
Pooled	Q	11.9	20.3	28.7	32.1	62.8	78.1	52.6	**58.2**
Model	MSEP	0.54	0.69	1.19	18.41	253.60	2364.91	24.58	24.03
	MSET	0.00012	0.00034	0.00038	0.00088	**0.0031**	0.0052	0.0018	0.0029

Table 6.1: Errors for all Feature sets within each period

Table 6.1 provides a comprehensive summary of the results derived from different feature sets as detailed in Section 5.4.4. All feature sets are trained with consistent model configurations for each period and are assessed based on three criteria: Quantile-loss (Q), Price MSE (MSEP), and Target MSE (MSET). In each period, the lowest error for each measure is highlighted. Appendix D presents the feature importance plots discussed in this section.

Upon reviewing Table 6.1, it is evident that no individual feature set consistently outperforms others across all periods. Almost every feature set demonstrates periods where they excel and instances where their performance is poor compared to other sets. This observation implies that each feature set contains valuable information, but there are periods where they introduce a significant amount of noise to the model. To gain a deeper understanding of this dynamic, we delve into the feature importance of each Feature set.

Feature set 1 incorporates Google search data associated with the keywords "war" and "natural gas," in addition to the base model features. As depicted in Table 6.1, this specific feature set performs slightly worse than the Base set for most periods, except for p_4 and p_8. Notably, during period p_3, the feature set experiences a significant increase in error compared to the base set. Upon inspecting the feature importance plot, we observe that Natural gas searches rank as the fourth most dominant feature, while War searches are comparatively lower. For period p_4, where the feature set outperforms the Base set across all measures, we note that Natural gas searches are less prominent, while War searches gain importance. This observation suggests that War searches might contribute value to the predictions, whereas Natural gas searches mainly introduce noise.

Feature set 2, which includes data on German storage, exhibits results with higher errors during period p_2 and p_3, compared to the Base set. While demonstrating slight improvements across multiple periods, this feature set excels in period p_7. After examining the feature importance plots, it becomes apparent that the "Full(%)" feature gains significance while remaining insignificant during inaccurate periods.

Feature set 3, incorporating energy data, demonstrates strong performance in period p_4, p_7, and p_8, compared to the Base set. However, it exhibits substantial errors in period p_1 and p_2. Upon analyzing the imprecise periods, we observe that in period p_1, Sun power and CO_2 spot are prominent, while in period p_2, Wind energy and CO_2 spot rank among the four most important features. In periods of good performance, Wind energy takes precedence. Therefore, drawing a definitive conclusion from this observation is challenging, but it is evident that the features added in Feature Set 3 often dominate the model, potentially causing some disruptions.

Feature Set 4 incorporates all features related to oil and coal prices. This feature set generally yields robust results, except for period p_3, which incurs higher errors than the Base set. Notably, the feature set excels in period p_5, characterized by volatile price fluctuations. Examining the model behavior, the features in this set provide value as they are consistently present in accurate periods while not being overly dominant in inaccurate periods.

Feature Set 5 investigates whether incorporating stock data for four major companies related to the natural gas supply can offer additional insights into the model. As shown in Table 6.1, this feature set performs similarly to the Base set, with some improvements and some declines in performance. Examining the importance of features, stock price features generally gain significance in most periods, with instances where they dominate other features. The stocks of these companies are highly correlated with commodity prices, potentially leading the model to assign higher importance to these features. Despite offering some value, especially in period p_6, they may introduce significant noise.

The Full model encompasses all available features. As depicted in Table 6.1, the results unveil a volatile pattern in the performance. Notably, period p_1, p_2, and p_4 present highly favorable predictions in comparison to alternative feature sets, whereas period p_7 and p_8 exhibit substantial increases in error. These observations suggest that adding extra features brings considerable value in specific periods but also introduces noise into the model. This noise becomes particularly apparent in period p_7 and p_8. During these periods, we observe dominant behavior of the features Natural gas searches, Full(%), CO_2 spot, Shipping price, and EOAN. These findings imply that identifying an optimal feature set for the model can significantly mitigate loss during testing.

The final feature set examined is the Pooled model, which incorporates price data from

each natural gas hub consolidated into one dataset. Employing the static enrichment layers of the TFT model, this approach allows the model to generate separate predictions for each hub while concurrently capturing shared patterns. As indicated in Table 6.1, the Pooled model performs similarly to the Simple model, with specific improvements, particularly in the most recent periods.

In light of the uncertain trajectory of gas prices in the future, our objective is to pinpoint a set of features that proves resilient across diverse periods, thereby ensuring adaptability to evolving dynamics in upcoming scenarios. Given limited computational power, employing sophisticated feature selection techniques is deemed too resource-intensive. To address the challenge of selecting the optimal subset of features, we leverage the transparency inherent in the model. Through an iterative process of removing and analyzing dominant features in periods with inaccurate predictions, we derive a final subset of features that, on average, outperforms other feature sets presented in Table 6.1. The final feature set is presented in Appendix C.

6.2 Post-Model Analysis

We conduct a comprehensive post-model analysis of the price forecasts for each period outlined in Section 5.3. Initially, we evaluate each model based on the criteria specified in Section 5.5. Subsequently, we delve into the identification of the most essential features. Finally, an analysis aimed at understanding the reasons behind the varying significance of certain features is presented.

6.2.1 Period 1

	Price MSE	Target MSE	Correct Way
TFT	0.31	0.00012	87%
ARIMA	0.57	0.00013	65%
XGBoost	0.63	0.00012	32%

Table 6.2: Performance evaluation for period p_1

The outcomes for the period p_1 are presented in Table 6.2. During this period, TFT demonstrated superior performance compared to ARIMA and XGBoost in terms of Price MSE and Correct way, while XGBoost exhibited similar performance in Target MSE. The Diebold-Mariano test in Appendix E.1 reveals that the TFT model significantly outperforms the other models concerning Price MSE and outperforms ARIMA in terms of Target MSE. As discussed in Section 5.3, period p_1 is characterized by a descending trend with a low standard deviation. The combination of this low standard deviation and the observation that 88 % of all prediction intervals fall within the price range contributes to the overall low error across all models in this period.

Figure 6.1 illustrates the model's assessment of feature importance. Notably, the model identifies the Temperature feature as significant during this timeframe. The average temperatures for February, March, and April across the entire dataset are 3.49, 6.42, and 10.0, respectively. Comparatively, the average temperatures in 2014 were 5.88, 8.46, and 12.2, indicating a hot season. This warmer climate could lead to a decreased demand for gas in terms of heating requirements, which may contribute to the observed decline in gas prices during this period.

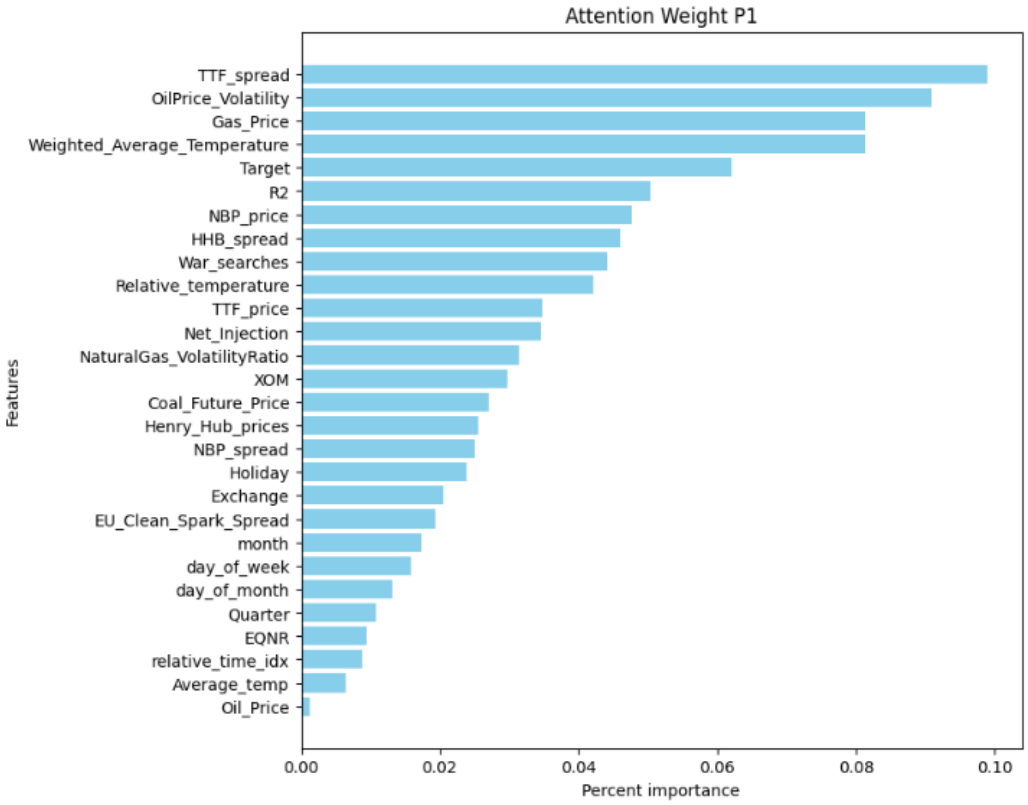


Figure 6.1: Attention weights p_1

As displayed in Figure 6.1 we observe that the TTF Spread feature is the most critical feature for the model attention during this period. Figure 6.2 illustrates this feature from January 2012 to December 2014, with the period p_1 colored orange. Notably, the TTF spread exhibits a meager and sustained value during this timeframe. The model acknowledges that such deviations tend to swiftly correct themselves, establishing this feature's crucial role in predicting future prices during this period.

In addition to the mentioned features, the model focuses on the Oil volatility and past values of the German gas prices.

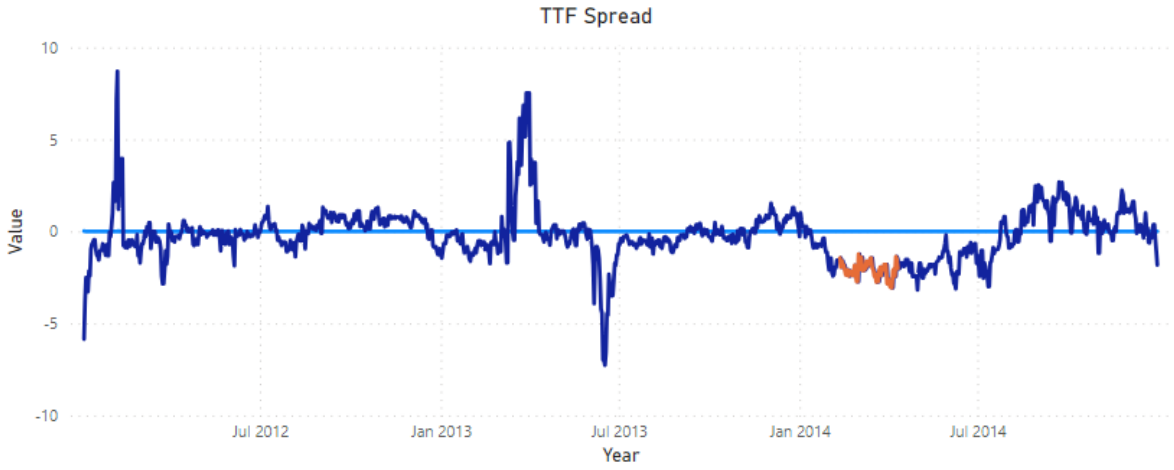


Figure 6.2: TTF Spread during period p_1

Source: Eikon

6.2.2 Period 2

	Price MSE	Target MSE	Correct Way
TFT	0.49	0.00033	72%
ARIMA	0.75	0.00036	61%
XGBoost	0.70	0.00033	70%

Table 6.3: Performance evaluation for period p_2

The period p_2 results are presented in Table 6.3. Although XGBoost and TFT show similar outcomes for Target MSE, TFT outperforms the other models regarding Price MSE. The Diebold-Mariano test results, however, indicate that this outperformance in MSE is not statistically significant. One possible explanation is that the TFT predictions consistently display a downward trend for all periods. This tendency is also reflected in the correctness measure, aligning with the number of declining periods. While TFT may demonstrate significantly lower MSE in falling forecasts, it may exhibit less flexibility in capturing rising prices.

Like period p_1, period p_2 exhibits a low standard deviation and a negative price trend. Despite having a lower standard deviation compared to period p_1, predicting this period is more challenging, as indicated by the results in Table 6.3. The complexity arises from the intricate nature of the price movements. In March 2019, there was a sharp price decline, followed by a steep increase in early April, followed by a gradual decrease. These dynamic shifts pose a challenge for the models, demanding a high degree of flexibility to capture the evolving patterns accurately

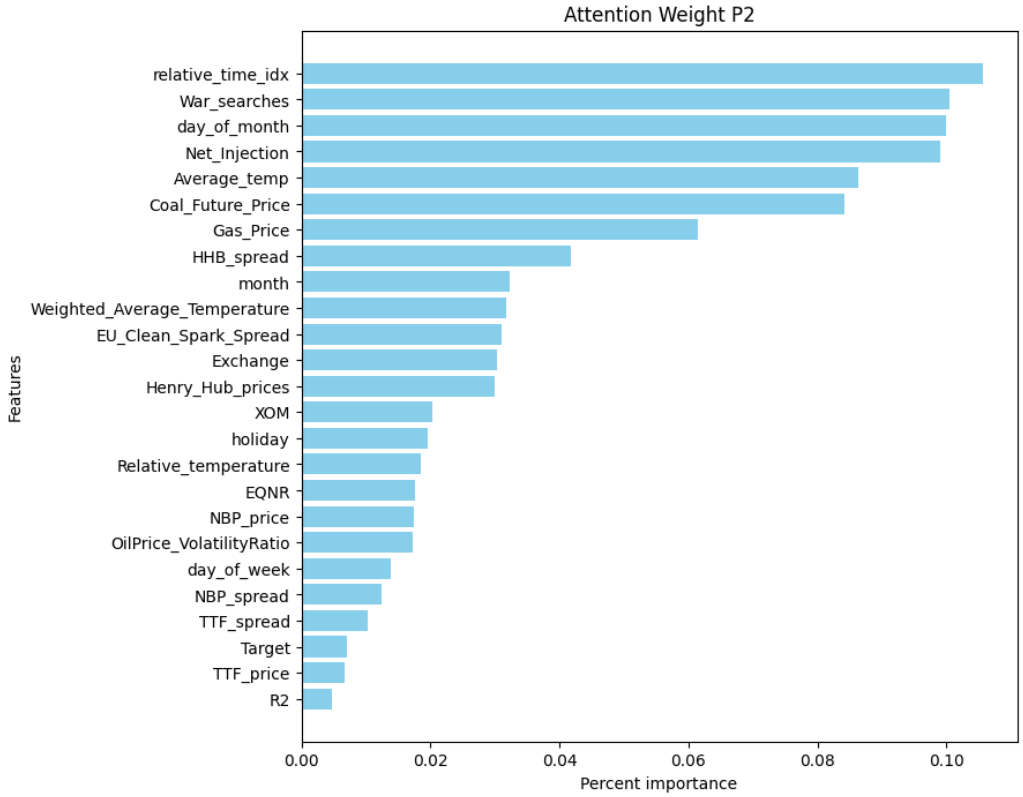


Figure 6.3: Attention weights p_2

In period p_2, the dominant features are the Relative time index, War searches, and Day of the month. First, the Relative time index, portraying the number of days since the starting day, provides temporal context, offering insights into evolving patterns over time. Similarly, the day-of-month feature, indicating the day in the current, introduces a temporal element that could capture seasonality effects and cyclical trends influencing natural gas prices. Finally, Google Data serves as an indicator for the volume of searches related to "War." Notably, searches for "War" may signify geopolitical tensions or conflicts, and the resultant geopolitical instability can impact energy markets, thereby influencing natural gas prices.

When examining Google searches for "War" displayed in Figure 6.4, an explanation for this feature's importance emerges. A pronounced spike was observed in April 2018, suggesting an influential event. Our primary hypothesis points to the Douma chemical attack in Syria during that month. In the aftermath of this attack, France, the UK, and the US conducted missile strikes on Syrian targets, escalating global tensions [Deutsch, 2023]. Despite the occurrence of this spike approximately a year before the current prediction period, it remains within the model's contextual window. The model may identify that the prediction period differs substantially from the previous year due to the noteworthy event in 2018.

As the model cannot rely on the historical price behavior from past years, it intensifies its attention on other seasonal features like net injection and average temperature. This involves identifying analogous seasonal periods beyond the same day of the previous year relative to the prediction period.

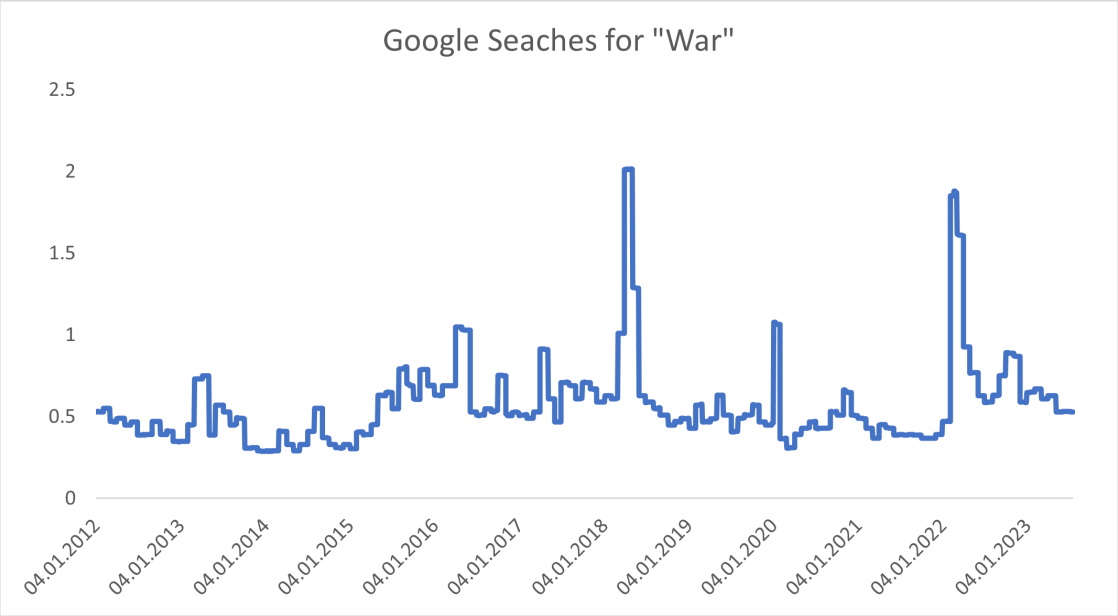


Figure 6.4: Google searches for "War"

Source: Google Trends

6.2.3 Period 3

	Price MSE	Target MSE	Correct Way
TFT	0.69	0.00033	83%
ARIMA	1.07	0.00035	65%
XGBoost	1.11	0.00033	52%

Table 6.4: Performance evaluation for period p_3

The period p_3 results are presented in Table 6.4. While ARIMA and XGBoost showcase similar outcomes, TFT outperforms the benchmark models, notably regarding Price MSE and the Correct way metric. The significance of this outperformance is further evident in Appendix E.1. Additionally, the TFT model significantly outperforms the ARIMA model in Target MSE, as detailed in Appendix E.2.

As indicated in Table 5.1, period p_3 witnesses an increase in gas prices, with a higher standard deviation compared to periods p_1 and p_2. Despite the higher standard deviation in this period, the model excels in regards to the correct way metric. This could be attributed to a consistent pattern in the data, characterized by a consistent upward price trend.

In period p_3, some important features observed from figure 6.5 are Relative temperature, Weighted average temperature, and Month. The Relative temperature and Weighted average temperature features hold substantial importance, potentially reflecting seasonal demand patterns and responding to heating and cooling needs. During February, March, and April in the prediction period p_3, the weather exhibits high volatility, fluctuating between icy and warm conditions. Specifically, the temperature was notably colder in April, registering 3 degrees below the average for this month. This could have contributed

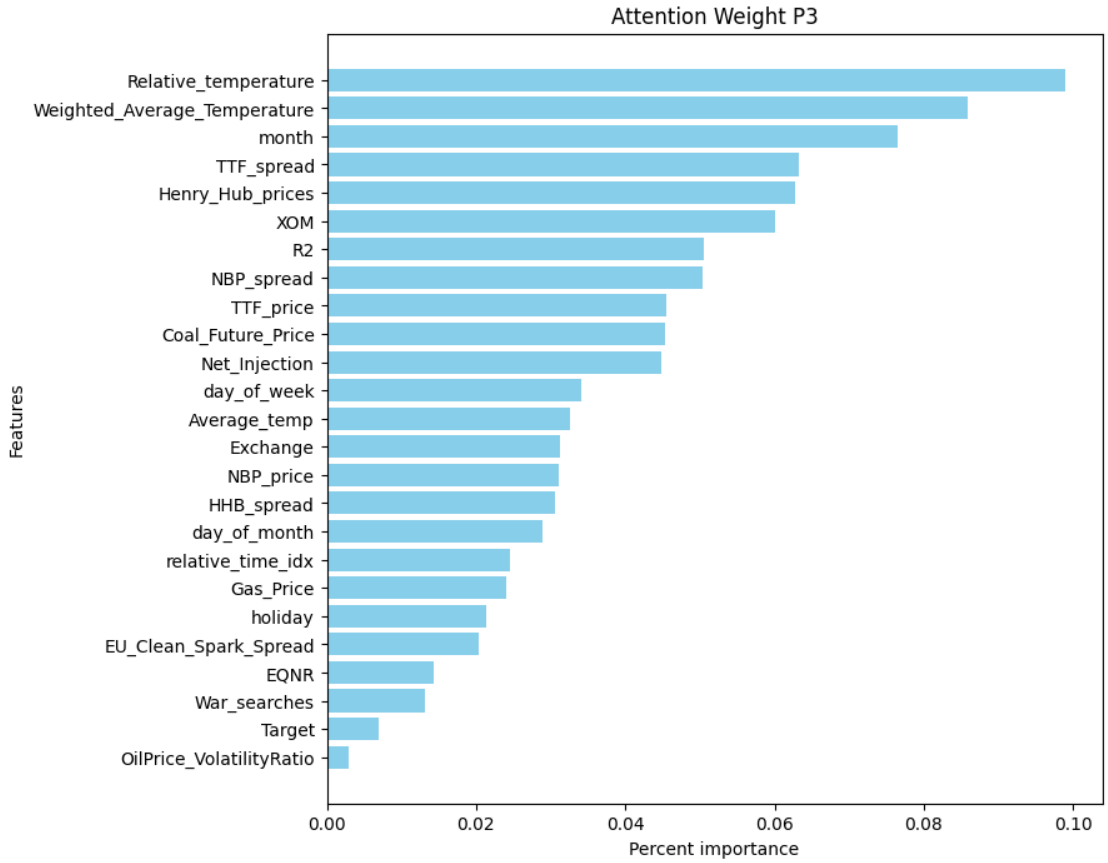


Figure 6.5: Attention weights p_3

to the observed increase in gas prices during this period.

In addition to the significance of the TFT spread as a feature during this period, Figure 6.5 also emphasizes the importance of Henry Hub prices. Figure 6.6 displays Henry Hub prices from January 2021 to July 2021, with values within the p_3 highlighted in blue. It is evident that just before this period, prices experienced a substantial spike, increasing by over 712 % from February 9th to February 17th. According to IEA [2021], this spike resulted from an extreme winter storm in parts of the United States, causing massive demand for heating. The resultant shortage in storage led to a decline in the global supply of American natural gas. This decrease could also contribute to the rise in German natural gas prices during this period.

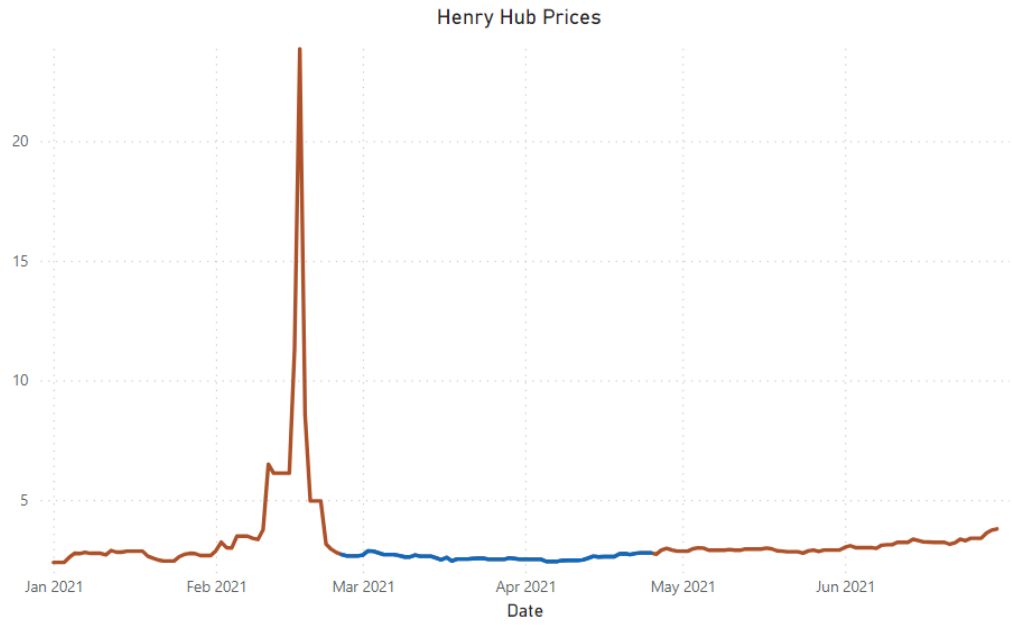


Figure 6.6: Henry Hub prices during period p_3

Source: Eikon

6.2.4 Period 4

	Price MSE	Target MSE	Correct Way
TFT	10.66	0.00085	83%
ARIMA	19.98	0.00097	63%
XGBoost	16.78	0.00086	63%

Table 6.5: Performance evaluation for period p_4

Analyzing the findings in Table 6.5, it is apparent that during period p_4, both machine learning models outperform ARIMA in Price MSE and Target MSE. TFT exhibits superior performance in Price MSE and price movement prediction. However, despite the considerable reduction in Price MSE compared to the other models, the results of the Diebold-Mariano test suggest that this improvement is not statistically significant at the 5% threshold.

The noticeable increase in Target and Price MSE for all models can be attributed to the rise in standard deviation for gas prices during period p_4. As indicated in Table 5.1, period p_4 exhibits a standard deviation of 6.15 with a net increase of 22.14, making minor prediction errors resulting in substantial errors in both Price and Target MSE.

In period p_4, visualized in Figure 6.7, the dominant feature is the Oil price volatility ratio. As Section 2.6 outlines, oil prices might affect German natural gas prices. The average correlation between oil and gas prices across all periods is 0.36. However, in period p_4 specifically, the correlation level spikes to 0.85, making oil prices an influential factor. Figure 6.8 illustrates the Oil price volatility from March 2021 to December 2021, with period p_4 highlighted in orange. Throughout this period, we observed significant spikes in the feature. These spikes, combined with an increase in the correlation between gas and coal, might explain the significance of the oil price volatility in this period.

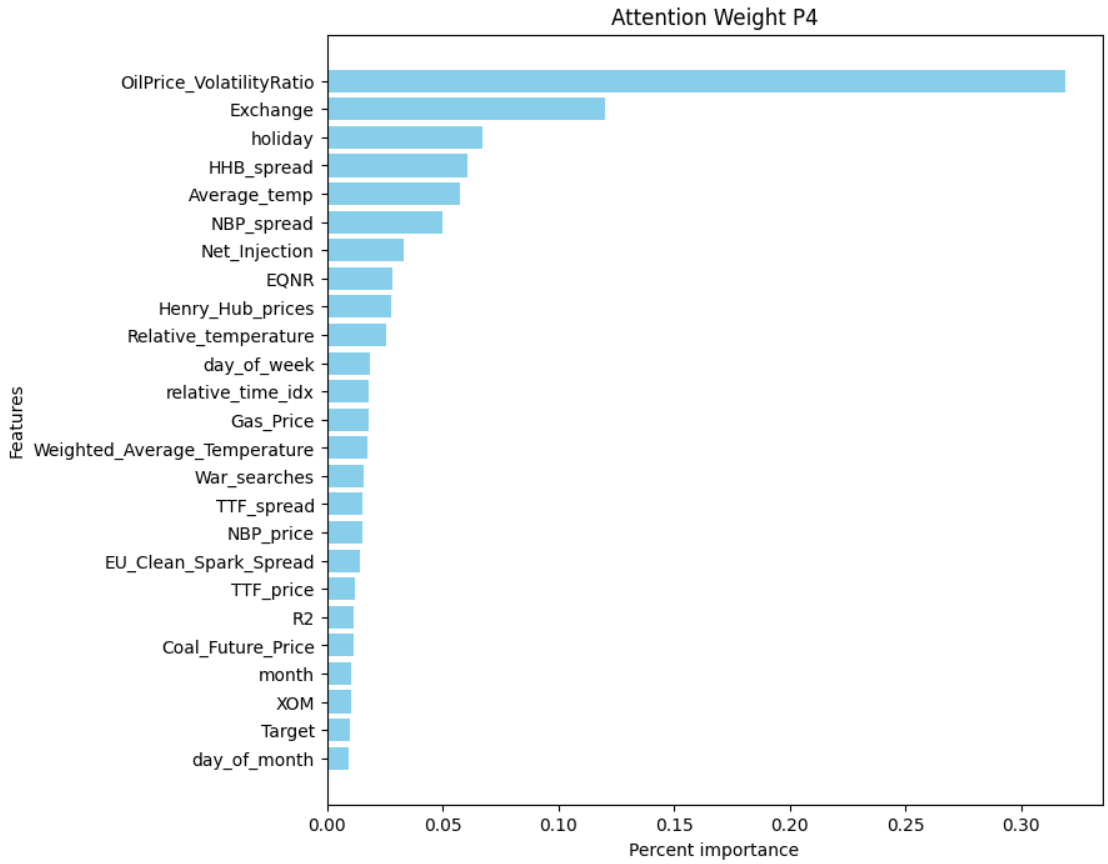


Figure 6.7: Attention weights p_4

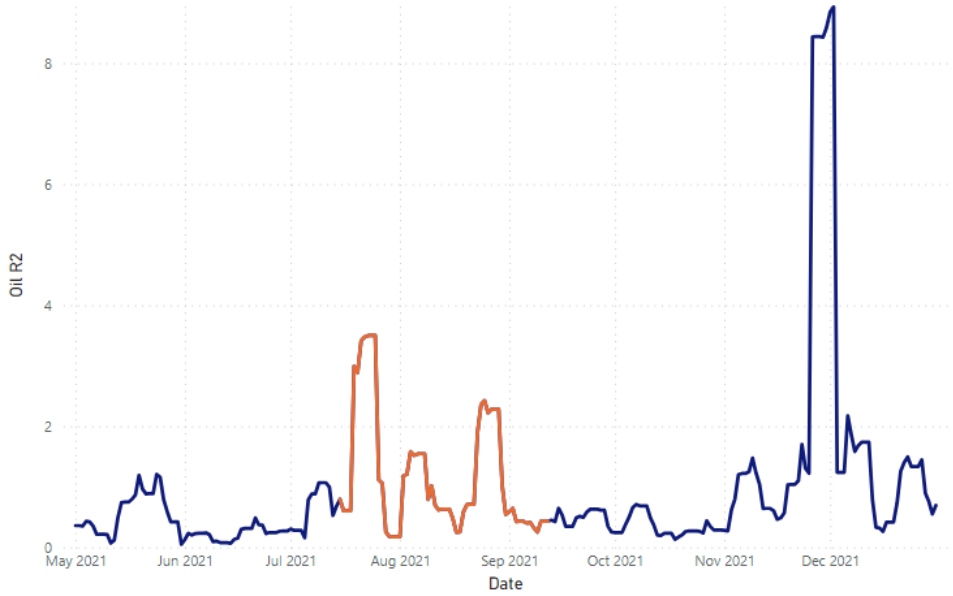


Figure 6.8: Oil Price volatility p_4

Source: Eikon

6.2.5 Period 5

	Price MSE	Target MSE	Correct Way
TFT	248.95	0.0032	65%
ARIMA	273.74	0.0035	47%
XGBoost	255.25	0.0032	47%

Table 6.6: Performance evaluation for period p_5

The period p_5 results are detailed in Table 6.6. As depicted in Section 5.3, period p_5 signifies the onset of the two most extreme periods, a characteristic further underscored by the outcomes in Table 6.6. According to Table 5.1, this period is characterized by a standard deviation of 17.1 and a net price increase of 28.9. The predictions exhibit approximately equal proportions of rises and falls, underscoring the challenging nature of accurate predictions during this period. All models exhibit notably higher errors during this period, with TFT marginally outperforming others regarding Price MSE and the correct way metric. From Appendices E.1 and E.2, it is evident that the TFT model significantly outperforms ARIMA in terms of Price MSE; however, this is not in terms of Target MSE.

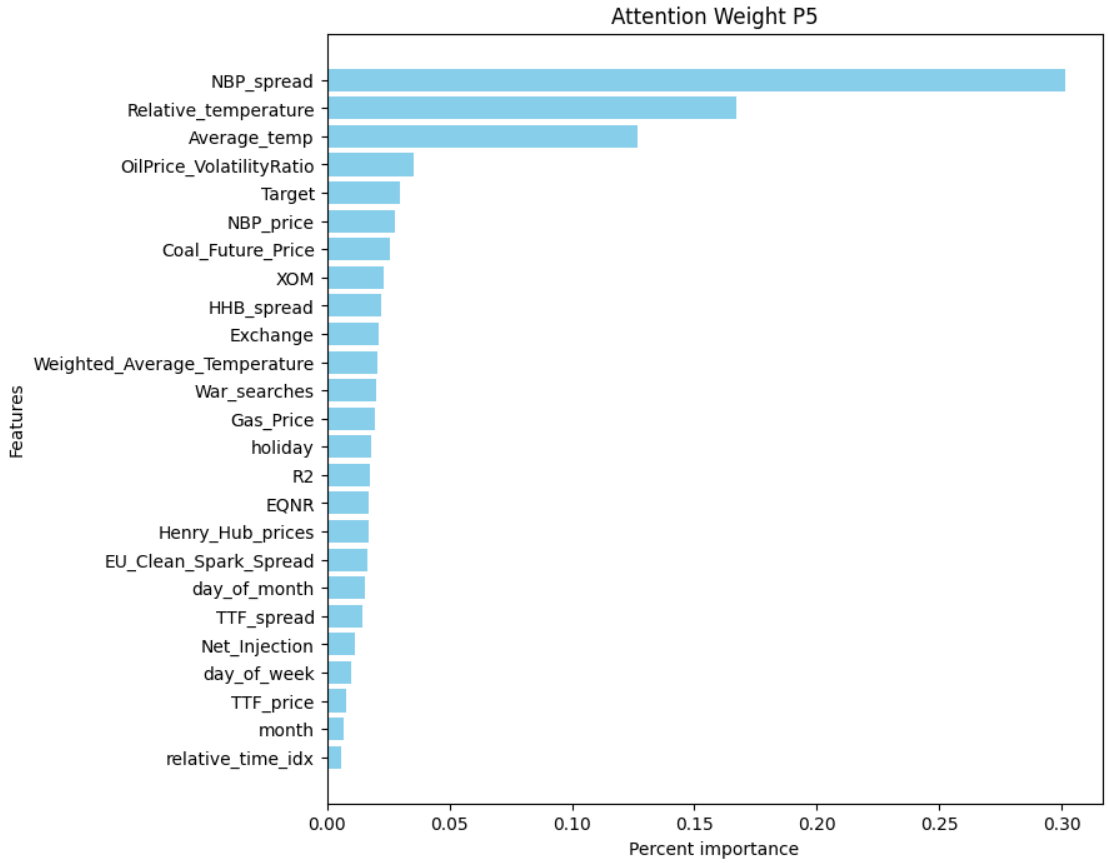


Figure 6.9: Attention weights p_5

In period p_5, we observe from Figure 6.9 that the feature NBP-spread emerges as the dominant factor along with temperature features. NBP spread indicates the price disparity between German and British natural gas, and its significance can be attributed to several reasons. Firstly, disparities in natural gas supply and demand between the UK and Germany can induce fluctuations in their respective prices. Changes in production, import/export volumes, or disruptions in the natural gas supply chain may contribute to these variations, with the observed price difference reflecting imbalances in supply and demand and enhancing the predictive efficacy of this variable. The NBP spread, historically averaging around 20, exhibited notable deviations with average values of 67.96 in 2021 and 80.79 in 2022. Figure 6.10 illustrates the behavior of the NBP spread in period p_5. NBP reached its all-time low at -65.29, indicating an unusual event in the market. The subsequent day saw a swift correction, with the value rebounding to 6.07. This intriguing pattern suggests that the NBP spread could serve as an indicator of potential future price fluctuations.

Moreover, the Relative temperature feature is an essential factor in this period. Notably, May 2022 experienced hot weather. Over the five years from 2018 to 2022, the average temperature for May stood at 13.77. However, in May 2022, an average temperature of 15.73 was recorded, signifying a substantial deviation towards higher temperatures. These elevated temperatures likely played a role in the downturn of gas prices, declining from approximately 100 to 80 during that month.

Figure 6.10: NBP Spread in p_5

Source: Eikon

6.2.6 Period 6

	Price MSE	Target MSE	Correct Way
TFT	1757.26	0.0054	65%
ARIMA	1967.23	0.0059	60%
XGBoost	1820.56	0.0052	58%

Table 6.7: Performance evaluation for period p_6

Table 5.1 reveals that period p_6 is the most extreme, characterized by a substantial standard deviation of 41.4 and a price range difference of 149 between the highest and lowest prices. The challenging nature of this period is further emphasized by the distribution of prediction directions, with 57% trending downward and 43% upward. This difficulty is reflected in the results presented in Table 6.7, where all models exhibit high errors. Among them, TFT demonstrates the lowest errors in terms of Price MSE, while XGBoost achieves the lowest errors in target MSE. From Appendix E, we observe that TFT outperforms the ARIMA model significantly in terms of Price MSE and Target MSE.

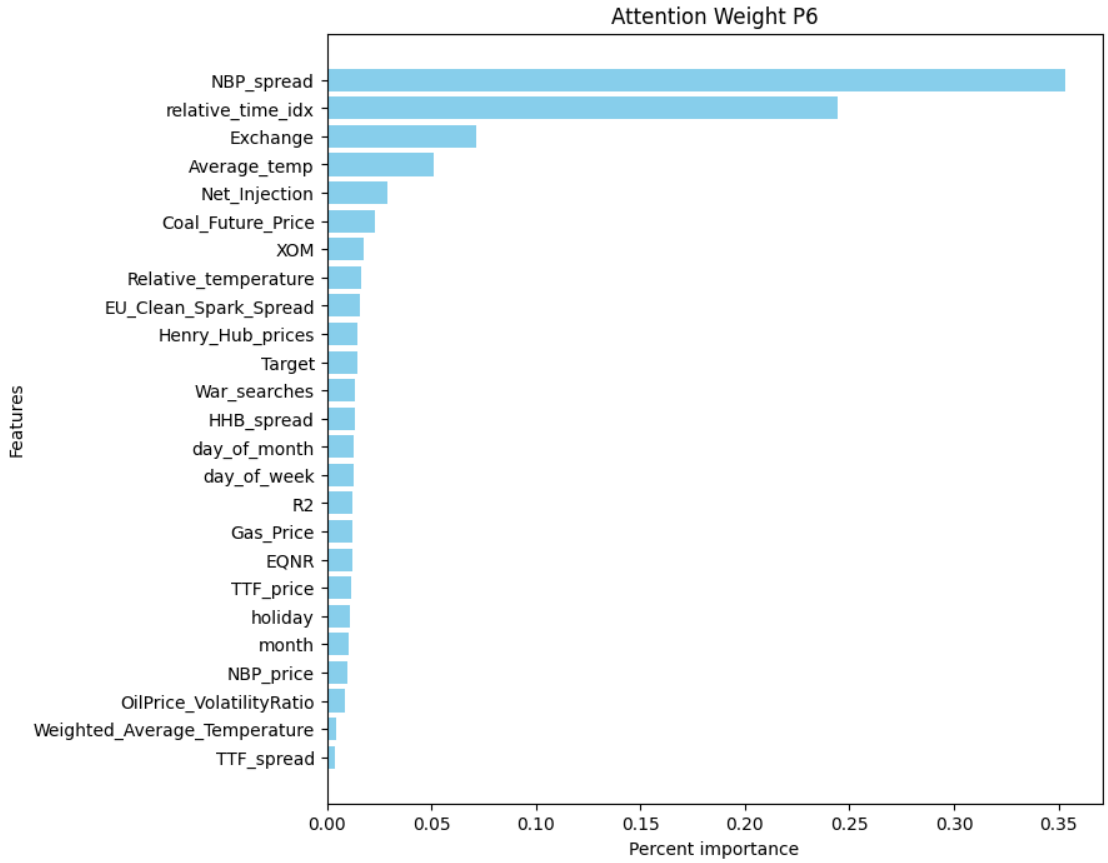


Figure 6.11: Attention weights p_6

Figure 6.11 highlights the most influential features of this period. The most important features are NBP spread, Relative time index, and Exchange rate. During period p_6, the volatile pattern observed in the NBP spread, illustrated in Figure 6.10, during period p_5, continues. This continuity in behavior could explain why this feature remains of high interest in this period.

In period p_6, there is a notable and rapidly changing pattern in gas prices, characterized by sharp increases and significant declines. The uncertainty surrounding future price patterns could be a key factor contributing to the significance of the Relative time index feature. The model's consideration of recent price data becomes particularly relevant in this scenario, as the absence of observed comparable data implies that recent trends may carry greater weight in predicting future price movements.

Figure 6.12 illustrates the exchange rates from 2019 to 2023, with the values corresponding to period p_6 highlighted in orange. Within this time frame, the exchange rate reaches its lowest value of the entire dataset, specifically on September 27th, with a recorded value of 0.96. This significant occurrence increases the model's focus on this particular feature.

Figure 6.12: Exchange rate in p_6

Source: Eikon

6.2.7 Period 7

	Price MSE	Target MSE	Correct Way
TFT	13.59	0.0015	93%
ARIMA	23.21	0.0017	60%
XGBoost	29.17	0.0016	35%

Table 6.8: Performance evaluation for period p_7

The outcomes for the period p_7 are presented in Table 6.8. Following the two extreme periods, this interval appears comparatively more stable and similar to period p_4, characterized by a negative price trend. TFT outperforms other models for all evaluation measures. The Diebold-Mariano test in Appendix E reveals a statistically significant outperformance for both measures compared to the benchmark models. However, it is noteworthy that TFT does not surpass XGBoost under the 5% threshold in Target MSE.

In period p_7, visualized in Figure 6.13, the dominant features are TTF price and NBP spread. The feature TTF price, however, emerges as the dominant factor.

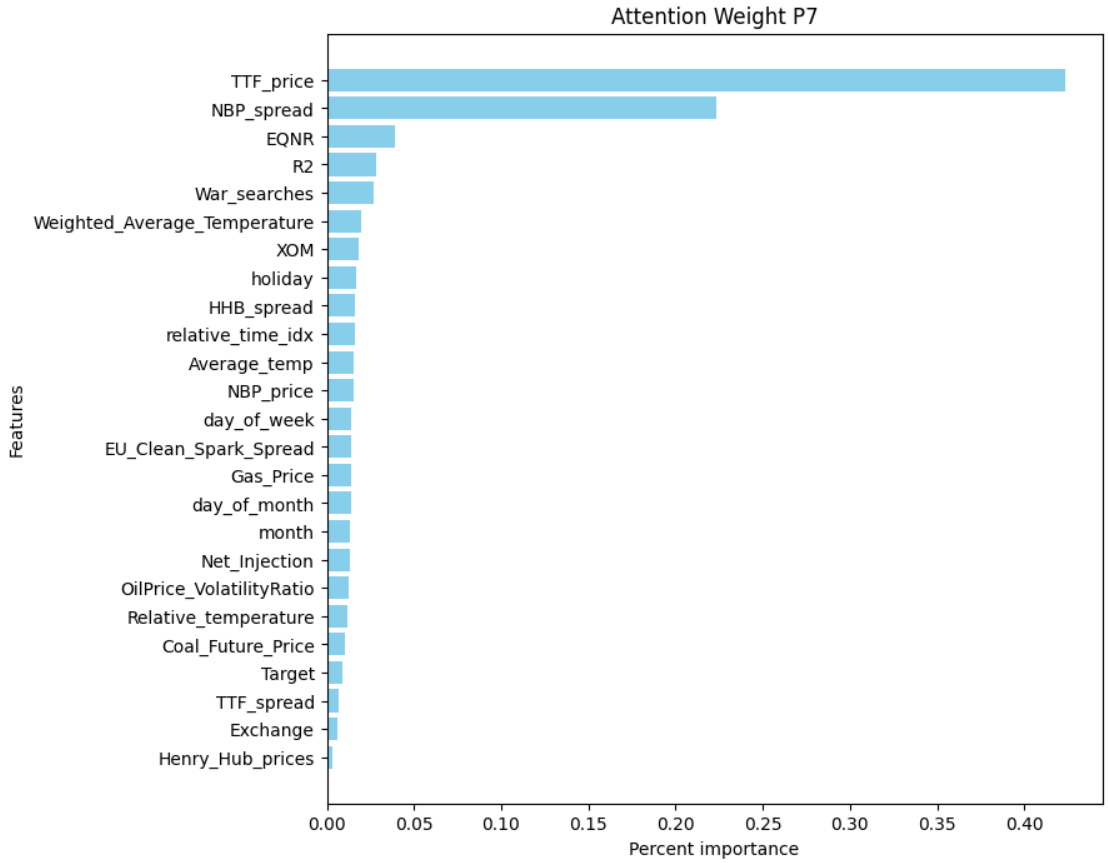


Figure 6.13: Attention weights p_7

Analyzing Figure 6.14, we observe that following an interval of substantial volatility in 2022, the beginning of 2023 marks a more stable phase in TTF prices. The model recognizes this deviation from the previous year's pattern. To delve deeper into the relationship between TTF and German gas prices, Figure 6.15 depicts the correlation development between the two gas hubs based on the past 365 days. In period p_7, the correlation regains significance after a deviation in 2022. The upswing in correlation and the absence of extreme volatility might explain why the model assigns high importance to TTF prices during this period.

Similar to the stabilization observed in the TTF price, the NPB spread also steadies during this period. These features could indicate a potential new regime in the gas market.

Figure 6.14: TTF Price from January 2022 to July 2023

Source: Eikon

Figure 6.15: Correlation between German Gas Prices and TTF Prices

Source: Eikon

6.2.8 Period 8

	Price MSE	Target MSE	Correct Way
TFT	18.77	0.0029	78%
ARIMA	23.64	0.0033	58%
XGBoost	23.13	0.0030	48%

Table 6.9: Performance evaluation for period p_8

As outlined in Section 2.5, natural gas prices have recently entered an unstable equilibrium, a notion supported by Table 5.1. In period p_8, a standard deviation of 4.6 is observed, with a slight negative trend. Table 6.9 presents the outcomes for this period. TFT surpasses the other benchmark models across all evaluation measures, although, as observed in Appendix E, the outperformance for the MSE measures is not statistically significant. This could be attributed to the observations in period p_8, where the TFT model lags in capturing the rapid decline in prices, only responding to the subsequent sharp increase. This lagging behavior might explain why the observed performance improvement is not statistically significant. This could be attributed to the fact that in period p_8, marked by a swift price decline succeeded by a sharp increase, the TFT model seems to delay capturing the downward trend before responding to the subsequent rise.

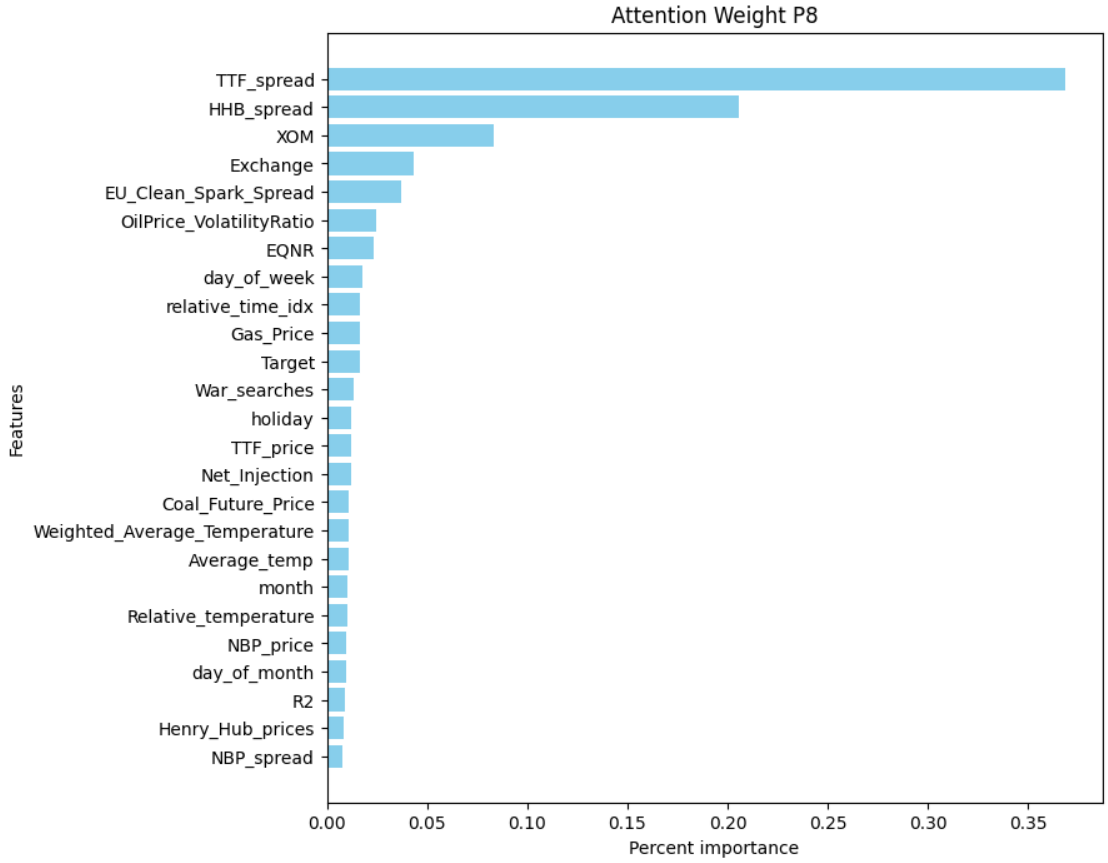


Figure 6.16: Attention weights p_8

In the last period, p_8, the predominant features are TTF spread and HHB spread, as seen in Figure 6.16. A closer examination of the HHB spread feature in Figure 6.17 reveals a stabilization pattern. After experiencing a significant deviation from its typical values in 2022, it gradually returns to a more normal state in 2023.

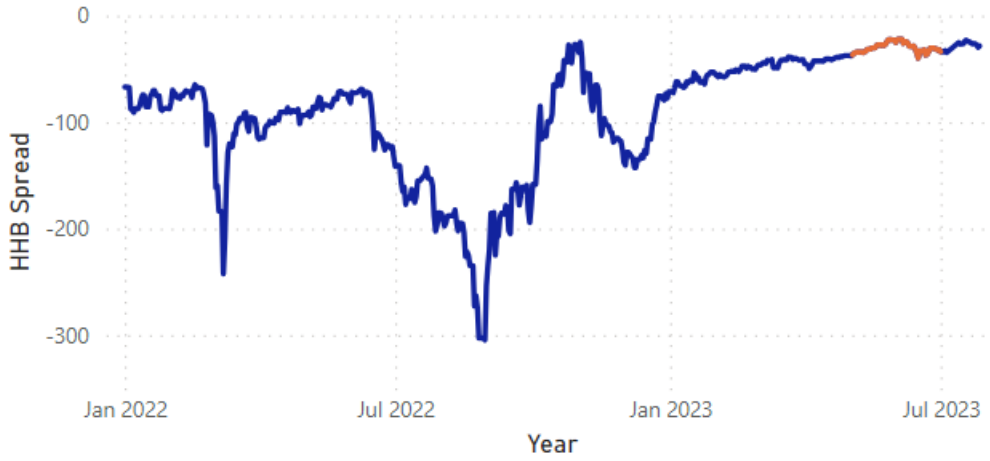


Figure 6.17: Henry Hub Spread from January 2022 to July 2023

Source: Eikon

In addition to stabilizing the hub spreads compared to the previous year, we can also observe from Figure 6.18 that the TTF spread is experiencing increased volatility during June 2023. These factors contribute to this feature being the most important during this period.

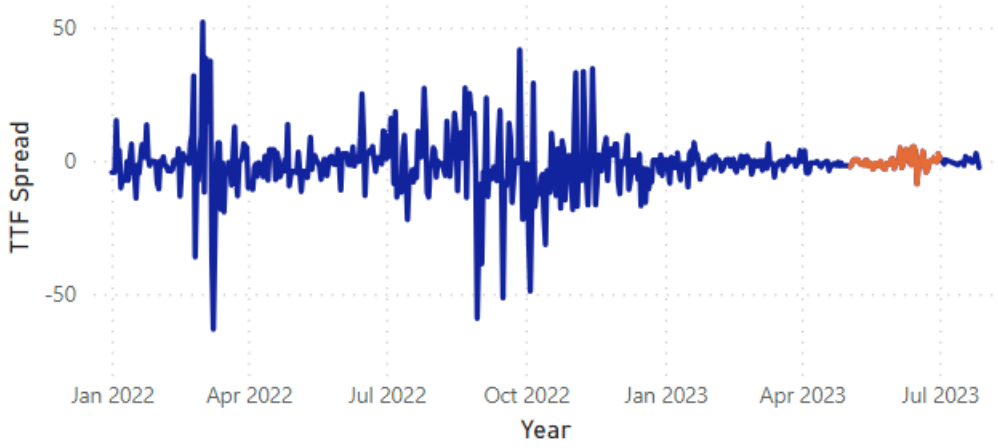


Figure 6.18: TTF Spread from January 2022 to July 2023

Source: Eikon

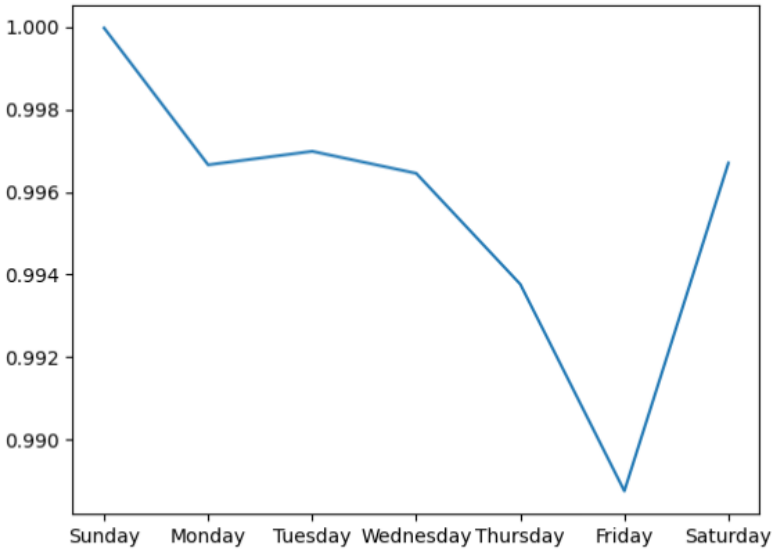
6.3 Temporal Variable analysis

To gain a more comprehensive understanding of the temporal dynamics of the model and test its pattern recognition capabilities, we conduct a sensitivity analysis focusing on temporal input features. These features encompass Day of the week, Month, and Day of the month, pivotal contributors to the predictive process of the TFT model, serving as known future inputs. The analysis is conducted by changing input values for each feature and observing the effect on predictions. It is crucial to emphasize that the analysis is conducted within the specific context of the final day of July 2023. This context will influence the overall predictions. However, our primary focus is on understanding the relative effects of the predictions rather than making absolute assessments. For this analysis, we utilize the model trained on the entire dataset.

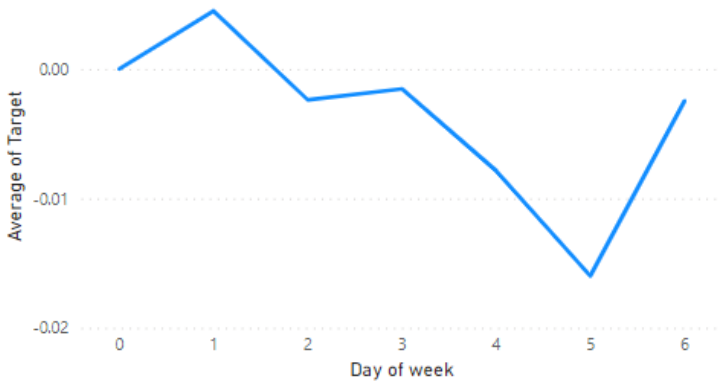
6.3.1 Day of the week

The Day of the Week feature ranges from Sunday, denoted 0, to Saturday, denoted 6. Figure 3a shows the TFT predictions for the scaled returns, while 3b illustrates the historical average return for each day. Scaled returns below 1 correspond to negative actual returns.

The TFT model consistently predicts negative values for all days except Sundays. Notably, it predicts Sundays as zero, ranking Mondays and Tuesdays as the highest and Fridays as the lowest. In comparison to Figure 3b, although predicting Tuesdays slightly higher than Mondays, the TFT model exhibits a pattern similar to the historically observed values. In their study, Meek & Hoelscher [2023] observed a significant positive impact on Mondays and Tuesdays and a significant adverse effect on Thursdays when analyzing Henry hub returns. The findings depicted in Figure 6.19 validate these results for German gas prices and reveal a negative impact on Fridays for the TFT model.



(a) TFT predictions

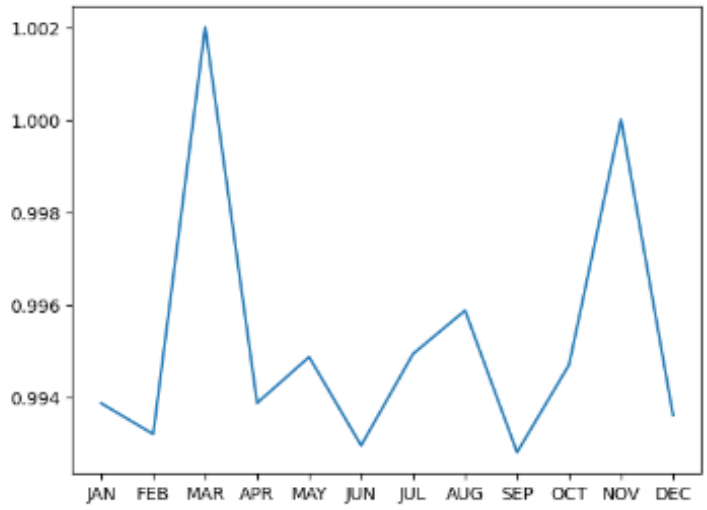


(b) Historically observed returns

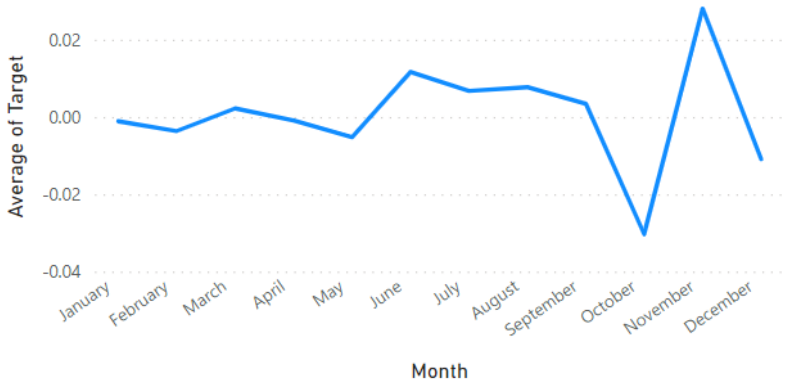
Figure 6.19: Day of the week effect

6.3.2 Month

Figure 6.20a presents the outcomes of the sensitivity analysis concerning the Month feature, while Figure 6.20b displays the observed historical average returns for each month. The TFT analysis reveals that the model predicts the highest values for March, August, and November and the lowest for February, June, and September. In comparison to the historical observed values, it is apparent that the November effect is noticeable. The model, however, tends to overpredict in March and underpredict in June. It is noteworthy that March and June exhibit some of the highest variance among the months, which could potentially make the model's predictions more sensitive to the context for these forecasts. This heightened sensitivity might contribute to the model's deviation from the historically observed values.



(a) TFT predictions

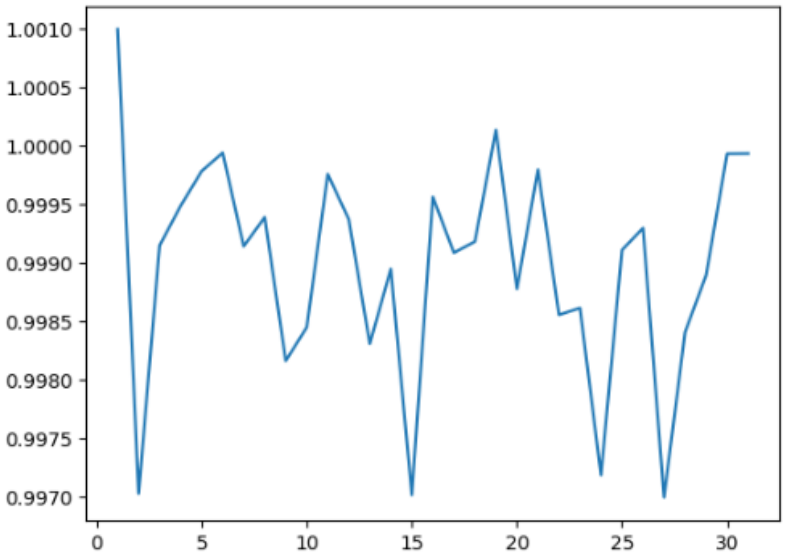


(b) Historically observed returns

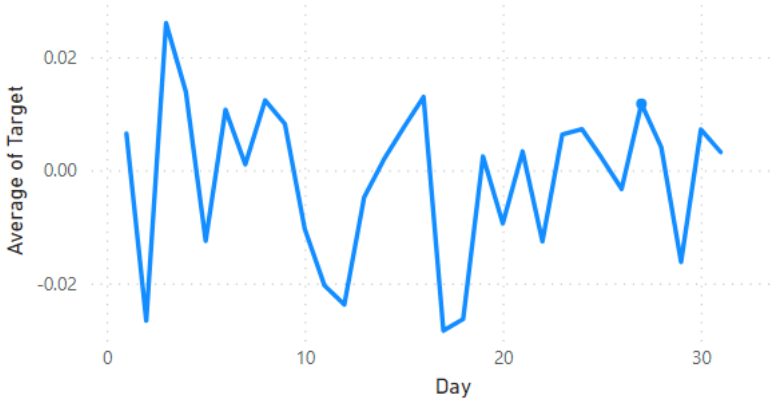
Figure 6.20: Month effect

6.3.3 Day of month

The last feature we analyze is the Day of the month feature. Figure 6.21a displays the sensitivity analysis results in the model, while Figure 6.21b presents the historical observed average returns. Upon comparing the figures, it becomes apparent that the TFT model accurately captures many of the recurrent dynamics observed historically. Notably, it reflects historical patterns at the beginning and end of the month, where the initial day typically exhibits higher returns, followed by a decline on the second day before experiencing an upward trend in days 3 and 4. Towards the end of the month, day 30 often demonstrates elevated values. Moreover, the model effectively captures many correct fluctuations throughout the month, although these fluctuations are a semblance of random. Qadan et al. [2019] identified a notable Turn-of-Month effect in the monthly Henry Hub prices, signifying higher returns around the transition between months. Figure 6.21a illustrates increased returns on days 1, 3, 4, 30, and 31, excluding day 2. This might indicate that the model captures some Turn-of-Month effect.



(a) TFT



(b) Historically observed returns

Figure 6.21: Day of month effect

6.3.4 Decoder Importance

To assess the significance of each feature, we examine the attention importance of the decoder. The results from the TFT model are depicted in Figure 6.22. Notably, the Day of the Week feature exhibits the highest significance, whereas the Day of the Month shows the most negligible significance. The significance difference becomes more apparent when examining the variation in TFT predictions for each feature, with the Day of the Week feature demonstrating the highest sensitivity to input values.

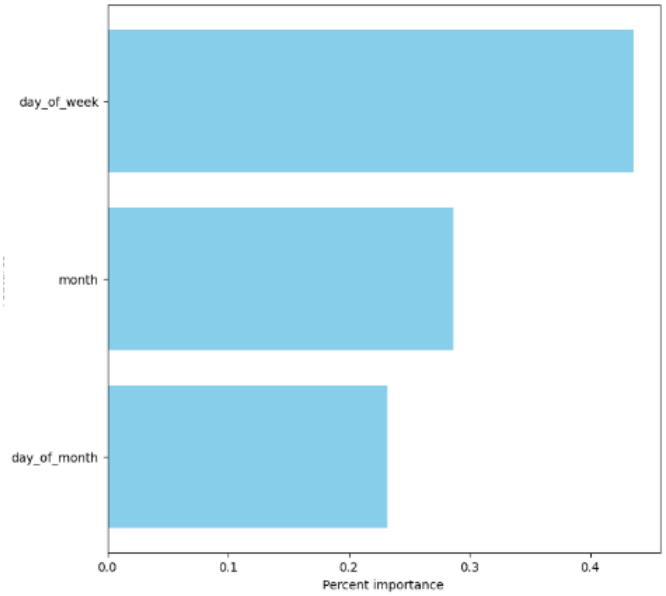


Figure 6.22: Decoder importance

Chapter 7

Conclusion

Our discoveries highlight various aspects of how the TFT model operates. First, we have highlighted the TFT model's ability to detect patterns in the data. Throughout all periods, the model has captured significant events in its contextual data, enabling it to predict future prices based on past behavior. Additionally, we observe that the model demonstrates an excellent ability to recognize patterns in temporal features, aligning with historical observations. This behavior has enabled the TFT model to consistently surpass both ARIMA and XGBoost regarding Price MSE, with statistically significant outperformance observed across numerous periods. Regarding Target MSE, the outcomes exhibit more balanced results than the XGBoost model, with only one period demonstrating significantly superior results. This observation might be attributed to the TFT model's capacity to discern trends in previous data and its ability to generate multi-horizon forecasts by considering past predictions and future known inputs. These aspects enable the TFT model to generate accurate sets of predictions rather than focusing solely on pointwise predictions. The same observation is apparent when analyzing the performance in terms of accurately predicting the directions of the price movements. The TFT model significantly outperforms the other models in most periods, underscoring the model's ability to identify recurring patterns in the data.

While the TFT model demonstrates commendable predictive accuracy under relatively normal circumstances, the model exhibits pronounced inaccuracies, particularly during extreme periods that have not been previously observed in the data. This becomes particularly evident in periods marked by increased market volatility, especially when compared to the XGBoost model, where the results are less significant. Upon analyzing the TFT model's behavior during increased market turbulence phases, we occasionally observe high attention on previous unrelated time steps. These time steps frequently display more stable behavior, causing the model to underpredict price fluctuations. This phenomenon becomes especially notable when the number of input features is increased, causing the TFT model to recognize patterns that might be unrelated to the current prediction period and display a delayed response to changing dynamics.

In addition to generating forecasts, the TFT proves its utility as a valuable tool for data analytics. Specifically, the TFT model offers insightful information regarding the significant features influencing forecasts. Our findings demonstrate that the most recurring features in terms of significance include spreads between other gas hubs, temperature data, and exchange rates. Gas hubs like the NBP and TTF are pivotal in forecasting German natural gas prices. These features, forming our Base set, align with the findings of Busse et al. [2012] from their research in 2012. While these features exhibit the most significance overall, relying solely on the base features does not consistently yield the best results across all periods. Incorporating additional features has proven to enhance the accuracy of our analysis. Some noteworthy discoveries regarding additional features include the impact of oil price volatility and Google searches for "war." The significance of the oil volatility feature suggests the potential presence of a spillover effect as described in Section 4.1. Despite previous studies indicating that the spillover effect between oil and gas prices has not been significant recently, the TFT model emphasizes that this feature still provides valuable information. The inclusion of Google searches is also intriguing. Integrating the search term for "war" provides the model with a signal of an unusual event occurring. This feature gains significance in some periods, mainly when the event occurred within its context window but not presently. This observation might indicate that this feature primarily indicates that previous unusual events have been concluded. We notice some significant spikes when analyzing the Google data in Figure 6.4. Still, it consistently maintains a low level, suggesting a lack of behavioral data for when these significant spikes occur. This is evident around the outbreak of the Ukraine-Russia war when the feature gains little significance. Obtaining additional data on the behavior of commodities during war outbreaks or analyzing different keywords that capture broader market disruptions could be interesting for further research.

The model's capability to discern patterns and events in the data introduces some additional complexities. As illustrated in Section 6.1, the model is highly responsive to its input data. Despite implementing sophisticated VSNs, the model continues to identify significant features that negatively influence its performance. This sensitivity to input features may be associated with a large context window of one year of data and a relatively small number of data points used for training. Including a single feature adds 365 data points for each context window. Considering that the model's learning period spans from a minimum of 800 to a maximum of 4000 days, it can identify patterns that lack value when applied out of the sample. Given the prolonged training time of the TFT model, implementing sophisticated feature selection algorithms becomes unfeasible due to the limitations in computational capacity. Our strategy of leveraging the model's transparency for feature selection has proven valuable. However, an extended data horizon could be beneficial to enhance the model's stability concerning its VSN. A more extensive utilization of additional data can be achieved by pooling data from various gas hubs. The concept of pooling, as demonstrated in Section 6.1, with price features only, should be extended in further research to include all features for each gas hub pooled into a single model. This approach could give the model a broader foundation to conclude, potentially leading to more stable VSNs.

A practical analysis of the financial value of the obtained predictions should be conducted to validate the results further. As discussed, the TFT model excels in precisely forecasting the direction of price movements, demonstrating remarkable accuracy even in periods of increased market volatility. This aspect could be of particular interest for assessment in a real-world application. One financial field in which this behavior might prove particularly advantageous lies within option trading. By leveraging the accuracy of the TFT model in predicting price trends, one can explore the potential to optimize entry and exit points in positions and enhance the ability to manage risks through more effective hedging strategies.

While our study has yielded promising results and valuable insights, it is essential to acknowledge certain limitations that open opportunities for further research and improvements. One primary constraint stems from the limitations in computational resources, which have restricted the range of values used in the hyperparameter tuning employed in our analysis. The intricate architecture of the TFT models results in substantial changes in behavior when specific parameters are adjusted. Identifying an optimal set of hyperparameters becomes crucial to enhance the performance of the TFT model. Future research should conduct a more comprehensive analysis of hyperparameter tuning and its dependence on the input set.

Another critical factor affecting the quality of our results is the constraint imposed by the availability of input data. Historical data regarding weather forecasts, electricity generation forecasts, and commodity spot data such as coal spot have not been available for our study. As described in Section 4.1, many of these factors have proven more significant for gas price forecasts than the proxy data utilized. Future research efforts could focus on securing the necessary resources or partnerships to access more valuable data, thus improving the overall robustness of the model.

In conclusion, while our current findings are promising, addressing the constraints related to computational resources, hyperparameter tuning, dataset size, and data quality presents exciting opportunities for future research. Overcoming these challenges could lead to a more nuanced and sophisticated understanding of the factors influencing our results, ultimately contributing to the advancement of the field and the applicability of our findings in practical settings.

Bibliography

A. Brown, S. P., & Yücel, M. K. (2008). What drives natural gas prices? *The Energy Journal*, *29*, 45–60. doi:10.5547/issn0195-6574-ej-vol29-no2-3.

Ackah, I. (2014). Determinants of natural gas demand in ghana. *Opec Energy Review*, *38*, 272–295. doi:http://dx.doi.org/10.1111/opec.12026.

Ali, A., Aliyuda, K., & Bello, A. (2021). Deep neural network model for improving price prediction of natural gas. *International Conference on Data Analytics for Business and Industry*, (pp. 113–117). doi:10.1109/ICDABI53623.2021.9655885.

Anderson, J. (2019). Record 2018 us natural gas demand largely driven by power demand: Ferc. *SP Global*, . URL: https://www.spglobal.com/commodityinsights/en/market-insights/latest-news/electric-power/041819-record-2018-us-natural-gas-demand-largely-driven-by-power-demand-ferc. Accessed: 2023-10-27.

Appun, K., Haas, Y., & Wettengel, J. (2023). Germany's energy consumption and power mix in charts? *Journalism for the energy transition*, . URL: https://www.cleanenergywire.org/factsheets/germanys-energy-consumption-and-power-mix-charts. Accessed: 2023-12-18.

Atil, A., Lahiani, A., & Nguyen, D. K. (2014). Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices. *Energy Policy*, *65*, 567–573. doi:10.1016/j.enpol.2013.09.064.

Bai, Y., & Okullo, S. J. (2023). Drivers and pass-through of the eu ets price: evidence from the power sector. *Energy Economics*, *123*. doi:https://doi.org/10.1016/j.eneco.2023.106698.

Balcombe, P., Brandon, N., & Giarola, S. (2018). Characterising the distribution of methane and carbon dioxide emissions from the natural gas supply chain. *Journal of Cleaner Production*, *172*, 2019–2032. doi:https://doi.org/10.1016/j.jclepro.2017.11.223.

Berrisch, J., & Florian, Z. (2022). Distributional modeling and forecasting of natural gas prices. *Journal of Forecasting*, *41*, 1065–1086. doi:https://doi.org/10.1002/for.2853.

Biresselioğlu, M., Yelkenci, T., & Oz, I. (2015). Investigating the natural gas supply security: a new perspective. *Energy*, *80*, 168–176. doi:https://doi.org/10.1016/j.energy.2014.11.060.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*, 307–327. doi:https://doi.org/10.1016/0304-4076(86)90063-1.

Boqiang, L., & Zhensheng, L. (2020). Can natural gas pricing reform establish an effective mechanism: Empirical evidence from china. *Energy Proceedings*, *12*. doi:10.46855/energy-proceedings-7426.

Box, E. P. G., & Jenkins, M. G. (2008). *Time Series Analysis*. (4th ed.). John Wiley Sons, Inc. doi:10.1002/9781118619193 ISBN: 9781118619193.

Van den Burg, G. J., & Williams, C. K. (2020). An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222*, . doi:https://doi.org/10.48550/arXiv.2003.06222.

Burke, P., & Yang, H. (2016). The price and income elasticities of natural gas demand: international evidence. *Energy Economics*, *59*, 466–474. doi:https://doi.org/10.1016/j.eneco.2016.08.025.

Busse, S., Helmholz, P., & Weinmann, M. (2012). Forecasting day ahead spot price movements of natural gas - an analysis of potential influence factors on basis of a narx neural network. *Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband der MKWI 2012*, *13*, 1395–1406. URL: https://www.researchgate.net/publication/233380926_Forecasting_day_ahead_spot_price_movements_of_natural_gas_-_An_analysis_of_potential_influence_factors_on_basis_of_a_NARX_neural_network.

Cai, Y., & Wu, Y. (2021). Understanding gas pricing mechanisms: Implications for the asian market. *Australasian Journal of Environmental Management*, *28*, 373–390. doi:10.1080/14486563.2021.1979111.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785–794). doi:https://doi.org/10.48550/arXiv.1603.02754.

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, . doi:https://doi.org/10.48550/arXiv.1511.07289.

Crow, D. J., Giarola, S., & Hawkes, A. D. (2018). A dynamic model of global natural gas supply. *Applied Energy*, *218*, 452–469. doi:https://doi.org/10.1016/j.apenergy.2018.02.182.

Dagher, L. (2012). Natural gas demand at the utility level: an application of dynamic elasticities. *Energy Economics*, *34*, 961–969. doi:https://doi.org/10.1016/j.eneco.2011.05.010.

De Bondt, W., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, *40*, 793–805. doi:https://doi.org/10.2307/2327804.

Deutsch, A. (2023). Chemical weapons watchdog blames syrian air force for douma attack. *Reuters*, . URL: https://www.reuters.com/world/middle-east/chemical-weapons-watchdog-blames-syrian-air-force-douma-attack-2023-01-27/. Accessed: 2023-12-20.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*, 427–431. doi:https://doi.org/10.1080/01621459.1979.10482531.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, *33*. doi:https://doi.org/10.1080/07350015.2014.983236.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*. doi:https://doi.org/10.1080/07350015.1995.10524599.

Dilaver, O., Dilaver, Z., & Hunt, L. (2014). What drives natural gas consumption in europe? analysis and projections. *Journal of Natural Gas Science and Engineering*, *19*, 125–136. doi:https://doi.org/10.1016/j.jngse.2014.04.002.

Dixit, A., & Jain, S. (2021). Effect of stationarity on traditional machine learning models: Time series analysis, . (pp. 303–308). doi:https://doi.org/10.1145/3474124.3474167.

Dixon, M. F., Halperin, I., & Bilokon, P. (2020). *Machine learning in finance*. (1st ed.). Springer. ISBN: 9783030410674.

Dong, K., Dong, X., & Sun, R. (2019). How did the price and income elasticities of natural gas demand in china evolve from 1999 to 2015? the role of natural gas price reform. *Petroleum Science*, *16*, 685–700. doi:https://doi.org/10.1007/s12182-019-0320-z.

Eckert, V. (2021). Germany launches nationwide gas trading hub the. *Reuters*, . URL: https://www.reuters.com/business/energy/germany-launches-nationwide-gas-trading-hub-2021-10-01/. Accessed: 2023-09-15.

Eckert, V., & Steitz, C. (2023). Explainer: Where does germany stand on gas supply? *Reuters*, . URL: https://www.reuters.com/business/energy/where-does-germany-stand-gas-supply-2023-01-20/. Accessed: 2023-10-26.

EIA (2021). In 2020, u.s. natural gas prices were the lowest in decades. *U.S. Energy Information Administration*, . URL: https://www.eia.gov/todayinenergy/detail.php?id=46376. Accessed: 2023-10-20.

EIA (2023). Natural gas prices fall in first half of 2023 amid record production and mild temperatures. *U.S. Energy Information Administration*, . URL: https://www.eia.gov/todayinenergy/detail.php?id=57200. Accessed: 2023-10-16.

Elbannan, M. A. (2014). The capital asset pricing model: An overview of the theory. *International Journal of Economics and Finance*, *10*, 216. doi:10.5539/ijef.v7n1p216.

Elias, R. S., Wahab, M. I. M., & Fang, L. (2016). The spark spread and clean spark spread option based valuation of a power plant with multiple turbines. *Energy Economics*, *59*, 314–327. doi:https://doi.org/10.1016/j.eneco.2016.08.003.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*, 179–211. doi:https://doi.org/10.1016/0364-0213(90)90002-E.

Energy Institute (2023). Statistical review of world energy, . URL: https://www.energyinst.org/statistical-review. Accessed: 2023-09-03.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, *50*, 987–1007. doi:https://doi.org/10.2307/1912773.

Erdoğdu, E. (2010). Natural gas demand in turkey. *Applied Energy*, *87*, 211–219. doi:https://doi.org/10.1016/j.apenergy.2009.07.006.

ESMA (2023). The august 2022 surge in the price of natural gas futures. *European Securities and Markets Authority*, . URL: https://www.esma.europa.eu/sites/default/files/2023-10/ESMA50-524821-2963_TRV_Article_the_August_2022_surge_in_the_price_of_natural_gas_futures.pdf. Accessed: 2023-10-01.

EU (2023). Diversification of gas supply sources and routes. *European Union*, . URL: https://energy.ec.europa.eu/topics/energy-security/diversification-gas-supply-sources-and-routes_en. Accessed: 2023-12-05.

Fama, E. (1991). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*, 383–417. doi:https://doi.org/10.2307/2325486.

Filis, G., Degiannakis, S., & Floros, C. (2011). Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis*, *20*, 152–164. doi:https://doi.org/10.1016/j.irfa.2011.02.014.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, *270*, 654–669. doi:https://doi.org/10.1016/j.ejor.2017.11.054.

Goor, H., & Scholtens, B. (2014). Modeling natural gas price volatility: The case of the uk gas market. *Energy*, *72*, 126–134. doi:https://doi.org/10.1016/j.energy.2014.05.016.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, *51*, 1–42. doi:https://doi.org/10.1145/3236009.

Halser, C., & Paraschiv, F. (2022). Pathways to overcoming natural gas dependency on russia—the german case. *Energies*, *15*, 4939. doi:10.3390/en15144939.

Hawkes, R. N., & Date, P. (2007). Medium-term horizon volatility forecasting: a comparative study. *Applied Stochastic Models in Business and Industry*, *23*, 465–481. doi:10.1002/asmb.684.

Heidari, S. (2020). How strategic behavior of natural gas exporters can affect the sectors of electricity, heating, and emission trading during the european energy transition. *Energies*, *13*, 5040. doi:10.3390/en13195040.

Ho, S. L., & Xie, M. (1998). The use of arima models for reliability forecasting and analysis. *Computers Industrial Engineering*, *35*, 213–216. doi:https://doi.org/10.1016/S0360-8352(98)00066-7.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*, 1735–1780. doi:https://doi.org/10.1162/neco.1997.9.8.1735.

Holland, B. (2021). 2020 average henry hub natural gas price hits lowest level in 25 years. *S&P Global*, . URL: https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/2020-average-henry-hub-natural-gas-price-hits-lowest-level-in-25-years-62023069. Accessed: 2023-11-01.

HSN (2022). The rise and fall of european gas prices. *Hellenic Shipping News*, . URL: https://www.hellenicshippingnews.com/the-rise-and-fall-of-european-gas-prices/. Accessed: 2023-09-30.

Hulshof, D., Van Der Maat, J.-P., & Mulder, M. (2016). Market fundamentals, competition and natural-gas prices. *Energy policy*, *94*, 480–491. doi:https://doi.org/10.1016/j.enpol.2015.12.016.

IEA (2019). Global energy review 2019. *International Energy Agency*, . URL: https://www.iea.org/reports/global-energy-review-2019. Accessed: 2023-10-19.

IEA (2021). Gas market report q4 2021 including global gas security review 2021. *Internatioal Energy Agency*, . URL: https://iea.blob.core.windows.net/assets/261043cc-0cb6-498b-98fa-a1f48715b91f/GasMarketReportQ42021.pdf. Accessed: 2023-12-20.

IEA (2022). Gas market report, q4-2022. *International Energy Agency*, . URL: https://iea.blob.core.windows.net/assets/5c108dc3-f19f-46c7-a157-f46f4172b75e/GasMarketReportQ42022.pdf. Accessed: 2023-11-05.

IGU (2023). Global gas report 2023. *International Gas Union*, . URL: https://www.igu.org/resources/global-gas-report-2023-edition/. Accessed: 2023-10-19.

Isidore, C. (2014). 89 straight days of lower gas prices. *Cable News Network (CNN)*, . URL: https://money.cnn.com/2014/12/23/news/economy/lower-gas-price-streak/. Accessed: 2023-10-25.

Jacobsen, S., & Abnett, K. (2022). Nord stream ruptures revealed europe grapples with gas plan. *Reuters*, . URL: https://www.reuters.com/business/energy/nord-stream-ruptures-revealed-europe-grapples-with-gas-plan-2022-10-18/. Accessed: 2023-10-23.

Kedzierski, M. (2023). Germany: how the gas sector changed in the crisis year of 2022. *Centre for Eastern European Studies Ośrodek Studiów Wschodnich*, . URL: https://www.osw.waw.pl/en/publikacje/osw-commentary/2023-01-12/germany-how-gas-sector-changed-crisis-year-2022. Accessed: 2023-12-10.

Keles, D., & Hasan, Y. (2020). Decarbonisation through coal phase-out in germany and europe—impact on emissions, electricity prices and power production. *Energy Policy*, *141*, 111472. doi:https://doi.org/10.1016/j.enpol.2020.111472.

Kemfert, C., Präger, F., Braunger, I., Hoffart, F. M., & Brauers, H. (2022). The expansion of natural gas infrastructure puts energy transitions at risk. *Nature Energy*, *7*, 582–587. doi:10.1038/s41560-022-01060-3.

Korosteleva, J. (2022). The implications of russia's invasion of ukraine for the eu energy market and businesses. *British Journal of Management*, *33*, 1678–1682. doi:https://doi.org/10.1111/1467-8551.12654.

Kotek, P., Selei, A., Borbála, T., & Felsmann, B. (2023). What can the eu do to address the high natural gas prices? *Energy Policy*, *173*. doi:https://doi.org/10.1016/j.enpol.2022.113312.

Kouchaksaraei, M., Movahedizadeh, H., & Mohammadalikhani, H. (2016). Determinant of the relationship between natural gas prices and leading natural gas countries' stock exchange. *International Journal of Economics and Finance*, *10*, 246. doi:10.5539/ijef.v8n4p246.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *NIPS'19: 33rd International Conference on Neural Information Processing Systems*, *471*, 5243–5253. doi:https://doi.org/10.48550/arXiv.1907.00235.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*, 1748–1764. doi:https://doi.org/10.1016/j.ijforecast.2021.03.012.

Lin, A. J., Chang, H. Y., & Hsiao, J. L. (2019). Does the baltic dry index drive volatility spillovers in the commodities, currency, or stock markets? *Transportation Research Part E: Logistics and Transportation Review*, *127*, 265–283. doi:https://doi.org/10.1016/j.tre.2019.05.013.

Lin, B., & Li, J. (2015). The spillover effects across natural gas and oil markets: Based on the vec–mgarch framework. *Applied Energy*, *155*, 229–241. doi:https://doi.org/10.1016/j.apenergy.2015.05.123.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*, 13–37. doi:https://doi.org/10.2307/1924119.

Madžarević, A., & Crnogorac, M. (2022). Security of supply as a major part of the energy security puzzle. *Energija Ekonomija Ekologija*, (pp. 28–37). doi:10.46793/EEE22-4.28M.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, *13*. doi:https://doi.org/10.1371/journal.pone.0194889.

Markowitz, H. M. (1952). Portfolio selection. *Journal of finance*, *7*, 77–91. doi:https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

Meek, A. C., & Hoelscher, S. A. (2023). Day-of-the-week effect: Petroleum and petroleum products. *Cogent Economics & Finance*, *11*, 2213876. doi:https://doi.org/10.1080/23322039.2023.2213876.

Moradi, A., Salehi, J., & Ravadanagh, S. N. (2022). Risk-based optimal decision-making strategy of a power-to-gas integrated energy-hub for exploitation arbitrage in day-ahead electricity and natural gas markets. *Sustainable Energy, Grids and Networks*, *31*, 100781. doi:https://doi.org/10.1016/j.segan.2022.100781.

Morgan, I. G. (1976). Stock prices and heteroscedasticity. *The Journal of Business*, *49*, 496–508. doi:10.1086/295881.

Morkunas, M., Černius, G., & Giriūnienė, G. (2019). Assessing business risks of natural gas trading companies: Evidence from get baltic. *Energies*, *12*, 2647. doi:10.3390/en12142647.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*, 768–783. doi:https://doi.org/10.2307/1910098.

Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics*, *29*, 46–63. doi:https://doi.org/10.1016/j.eneco.2006.04.003.

Nazir, A., Shaikh, A. K., Shah, A. S., & Khalil, A. (2023). Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (tft). *Results in Engineering*, *17*, 100888. doi:https://doi.org/10.1016/j.rineng.2023.100888.

Nick, S., & Thoenes, S. (2014). What drives natural gas prices? — a structural var approach. *Energy Economics*, *45*, 517–527. doi:https://doi.org/10.1016/j.eneco.2014.08.010.

NVIDIA (2023). Xgboost, . URL: https://www.nvidia.com/en-us/glossary/data-science/xgboost/. Accessed: 2023-12-19.

de Prado, M. L. (2019). Beyond econometrics: A roadmap towards financial machine learning. *Econometric Modeling: Theoretical Issues in Microeconometrics eJournal*, . doi:https://dx.doi.org/10.2139/ssrn.3365282.

Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, *214*, 242–268. doi:https://doi.org/10.1016/j.neucom.2016.06.014.

Prince, D. D., Marçal, E. F., & Pereira, P. L. V. (2022). Forecasting industrial production using its aggregated and disaggregated series or a combination of both: evidence from one emerging market economy. *Econometrics*, *10*, 27. doi:10.3390/econometrics10020027.

Qadan, M., Aharon, D. Y., & Eichel, R. (2019). Seasonal patterns and calendar anomalies in the commodity market for natural resources. *Resources Policy*, *63*, 101435. doi:https://doi.org/10.1016/j.resourpol.2019.101435.

Ram, M., Taklif, A., & F, A. (2019). Prediction of natural gas prices in european gas hubs using artificial neural network. *Petroleum Business Review*, (pp. 1–14). doi:10.22050/pbr.2019.113878.

Ray, P. P. (2023). Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. doi:https://doi.org/10.1016/j.iotcps.2023.04.003.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*, 386–408. doi:https://doi.org/10.1037/h0042519.

Rubaszek, M., & Szafranek, K. (2022). Have european natural gas prices decoupled from crude oil prices? evidence from tvp-var analysis. *Studies in Nonlinear Dynamics & Econometrics*, . doi:https://doi.org/10.1515/snde-2022-0051.

Scott, A. J., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, *30*, 507–512. doi:https://doi.org/10.2307/2529204.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*, 425–442. doi:https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Shelor, J. (2020). In year of extremes, 2019 natgas spot prices lowest in three years. *Natural Gas Intelligence*, . URL: https://www.naturalgasintel.com/in-year-of-extremes-2019-natgas-spot-prices-lowest-in-three-years/. Accessed: 2023-10-29.

Shi, X. (2016). Development of europe's gas hubs: Implications for east asia. *Natural gas industry B*, *3*, 357–366. doi:https://doi.org/10.1016/j.ngib.2016.11.001.

Shi, X., & H., V. (2016). Gas and lng trading hubs, hub indexation and destination flexibility in east asia. *Energy Policy*, *96*, 587–596. doi:10.1016/j.enpol.2016.06.032.

Shi, X., Shen, Y., & Wu, Y. (2019). Energy market financialization: Empirical evidence and implications from east asian lng markets. *Finance Research Letters*, *30*, 414–419. doi:10.1016/j.frl.2019.02.004.

Shumway, R. H., & Stoffer, D. S. (2017). Arima models. In *Time Series Analysis and Its Applications: With R Examples* (pp. 75–163). Cham: Springer International Publishing. doi:10.1007/978-3-319-52452-8_3 ISBN: 978-3-319-52452-8.

S&P (2021). Conversion base rates, . URL: https://www.spglobal.com/commodityinsights/en/our-methodology/conversion-tables. Accessed: 2023-09-1.

Speight, J. G. (2018). *Natural gas: a basic handbook*. (2nd ed.). Gulf Professional Publishing. ISBN: 9780128095706.

Stejskalová, J. (2023). Behavioral attention by google trends: Evidence from the car industry. *Ekonomicky Casopis/Journal of Economics*, *71*, 202–221.

Stocker, M., Baffes, J., & Vorisek, D. (2018). What triggered the oil price plunge of 2014-2016 and why it failed to deliver an economic impetus in eight charts. *World Bank Blogs*, . URL: https://blogs.worldbank.org/developmenttalk/what-triggered-oil-price-plunge-2014-2016-and-why-it-failed-deliver-economic-impetus-eight-charts. Accessed: 2023-11-14.

Su, M., Zhang, Z., Zhu, Y., Zha, D., & Wen, W. (2019). Data driven natural gas spot price prediction models using machine learning methods. *Energies*, *12*, 1680. doi:https://doi.org/10.3390/en12091680.

Tang, Y., Wang, Q., Xu, W., Wang, M., & Wang, Z. (2019). Natural gas price prediction with big data. *2019 IEEE International Conference on Big Data*, (pp. 5326–5330). doi:10.1109/BigData47090.2019.9006195.

Thara, D., PremaSudha, B., & Xiong, F. (2019). Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. *Pattern Recognition Letters*, *128*, 544–550. doi:https://doi.org/10.1016/j.patrec.2019.10.029.

Topal, M. O., Bas, A., & van Heerden, I. (2021). Exploring transformers in natural language generation: Gpt, bert, and xlnet. *arXiv preprint arXiv:2102.08036*, . doi:https://doi.org/10.48550/arXiv.2102.08036.

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*, 107299. doi:https://doi.org/10.1016/j.sigpro.2019.107299.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Neural Information Processing*, *30*. doi:https://doi.org/10.48550/arXiv.1706.03762.

Wang, K.-L., Fawson, C., Barrett, C., & McDonald, J. (2001). A flexible parametric garch model with an application to exchange rates. *Journal of Applied Econometrics*, *16*, 521–536. doi:https://doi.org/10.1002/jae.606.

Wang, T., Zhang, D., Ji, Q., & Shi, X. (2020). Market reforms and determinants of import natural gas prices in china. *Energy*, *196*, 117105. doi:10.1016/j.energy.2020.117105.

Wang, Y., Wang, Z., Kang, X., & Luo, Y. (2022). A novel interpretable model ensemble multivariate fast iterative filtering and temporal fusion transform for carbon price forecasting. *Energy Science and Engineering*, *11*, 1148–1179. doi:https://doi.org/10.1002/ese3.1380.

Wartenburg, E. (2023). Top 10: Germany's largest cities. *Deutsche Welle*, . URL: https://www.dw.com/en/top-10-germanys-largest-cities/g-52632011. Accessed: 2023-08-25.

Welle, D. (2022). Eu agrees to gas price cap, . URL: https://www.dw.com/en/eu-agrees-to-gas-price-cap/a-64154256. Accessed: 2023-12-02.

Xia, X., Wu, R., Liu, Y., Wu, J., & Lu, T. (2022). Value of long-term lng contracts: A theoretical and empirical study. *Frontiers in Earth Science*, *10*. doi:10.3389/feart.2022.1058592.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, *31*, 1235–1270. doi:https://doi.org/10.1162/neco_a_01199.

Zaid, M., & Khan, M. F. (2023). Russian-ukrainian war and its economic implications on the prices of strategic commodities. *Studies in Economics and Business Relations*, *3*, 1–16. doi:10.48185/sebr.v3i2.738.

Zhang, D., & Paltsev, S. (2016). The future of natural gas in china: effects of pricing reform and climate policy. *Climate Change Economics*, *7*, 1650012. doi:10.1142/S2010007816500123.

Zhang, D., Shi, M., & Shi, X. (2018). Oil indexation, market fundamentals, and natural gas prices: an investigation of the asian premium in natural gas trade. *Energy Economics*, *69*, 33–41. doi:10.1016/j.eneco.2017.11.001.

Zhang, M., Su, M., & Lu, W. (2015). An assessment of the security of china's natural gas supply system using two network models. *Energies*, *8*, 13710–13725. doi:https://doi.org/10.3390/en81212392.

Çelenk, H., & Bozlak, E. (2022). Hedge transactions, taxation, and accounting in the natural gas trade in the turkish energy sector. *Muhasebe Ve Vergi Uygulamaları Dergisi*, *15*, 155–181. doi:10.29067/muvu.937233.

Appendix

A Full dataset

Table 1 and 2 provide an extensive list of all data features utilized in this study with corresponding descriptions, data sources and frequency. The data sources utilized in this study include Eikon, Energy-Charts, Netztransparenz, Google Trends and Deutscher Wetterdienst. Refinitiv Eikon is the primary source of data and is a financial database providing access to a vast range of market data. Netztransparenz and Energy charts offer information pertaining to the German electricity market, while Deutscher Wetterdienst provides data on German weather.

Table 1: First part of the full feature set incorporated in the study

Feature	Description	From	Frequency
Exchange	EUR/USD Rate	Eikon	Daily
EU Clean Spark Spread	Calculation from section 4.1		Daily
Oil Price	Brent crude oil spot	Eikon	Daily
Wind Energy	Amount of wind energy generated in Germany	Netztransparenz	Daily
Average Temp.	Average temperature of the 10 largest cities in Germany	Deutscher Wetterdienst	Daily
Weighted Average Temp.	Calculation from section 4.1	Deutscher Wetterdienst	Daily
Fullness Rate	Natural gas storage levels in Germany	Eikon	Daily
Net Injection	Net storage Activity	Eikon	Daily
Shipping Price	Baltic Dry Index	Eikon	Daily
NaturalGas˙searches	Searches for "Natural gas" in Germany	Google Trends	Weekly
War searches	Searches for "war" globally	Google Trends	Weekly
TTF Price	Day ahead price of TTF Hub	Eikon	Daily
TTF Spread	Calculation from section 4.1		Daily
NBP Price	Day ahead price of NBP Hub in GBP	Eikon	Daily
NBP Spread	Calculation from section 4.1		Daily
Henry Hub Price	Day ahead price of Henry Hub	Eikon	Daily
Henry Hub spread	Calculation from section 4.1		Daily

Table 2: Second part of the full feature set incorporated in the study

Feature	Description	From	Frequency
Target	German Gas return		Daily
R2	Calculation from section 4.1		Daily
Relav R2	Calculation from section 4.1		Daily
Oil R2	R2 calculation on Oil Price		Daily
Relav Temperature	Calculation from section 4.1		Daily
Coal Futures Price	Rotterdam coal,1 month future price	Eikon	Daily
Gazprom	Gazprom stock price	Eikon	Daily
EOAN	E.ON AG stock price	Eikon	Daily
XOM	ExxonMobil stock price	Eikon	Daily
EQNR	Equinor stock price	Eikon	Daily
Gas Price	German gas Prices	Eikon	Daily
Sun Power	Amount of sun energy generated in Germany	Netztransparenz	Daily
Electricity Price	German day ahead electricity price	Energy-Charts	Daily
CO2 spot	Day ahead price for German $C0_2$ emission certificate	Energy-Charts	Daily
Day of week	Day of week from the Date		Daily
Month	Month from the Date		Daily
Day of month	Day of the month from the Date		Daily
Holiday	1 if the current day is a national holiday	Eikon	Daily
Relative time index	Relative position of input data with respect to the first prediction		Daily

B Model Parameters

B.1 ARIMA

Figure 1 presents a graphical representation of the summary statistics for the ARIMA model. The model is deployed using the Python library statsmodels version 0.14.1 Through the minimization of Akaike's Information Criterion, we identified the optimal ARIMA(3,1,3) model. The summary statistics reveal the significance of all coefficients, and the Ljung-Box test does not reject the null hypothesis. Notably, the White test is rejected, suggesting the presence of heteroscedasticity in the residuals.

SARIMAX Results

Dep. Variable:	y	No. Observations:	4228
Model:	ARIMA(3, 1, 3)	Log Likelihood	-11287.732
Date:	Sat, 16 Dec 2023	AIC	22589.465
Time:	01:20:37	BIC	22633.909
Sample:	0	HQIC	22605.175
	- 4228		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4488	0.019	23.195	0.000	0.411	0.487
ar.L2	-0.6889	0.015	-45.770	0.000	-0.718	-0.659
ar.L3	-0.3669	0.019	-19.512	0.000	-0.404	-0.330
ma.L1	-0.3031	0.019	-16.037	0.000	-0.340	-0.266
ma.L2	0.5635	0.014	40.800	0.000	0.536	0.591
ma.L3	0.5069	0.018	28.562	0.000	0.472	0.542
sigma2	12.2043	0.045	272.727	0.000	12.117	12.292

Ljung-Box (L1) (Q):	0.10	Jarque-Bera (JB):	1091065.92
Prob(Q):	0.75	Prob(JB):	0.00
Heteroskedasticity (H):	251.89	Skew:	0.35
Prob(H) (two-sided):	0.00	Kurtosis:	81.70

Figure 1: Summary statistics of the ARIMA model

B.2 XGBoost

Table 3 exhibits the hyperparameters employed in the XGBoost model deployed from the Python library xgboost. Additional parameters that were tested but remained unchanged include min˙child˙weight (ranging from 0 to 10), subsample (ranging from 0.5 to 0.8), and gamma (ranging from 1 to 3).

Library	xgboost	version 2.0.3
General Parameters	booster	”gbtree”
	objective	”reg:squarederror”
	epochs	100
	n˙estimators	100
	Learning rate	0.001
Parameters for Tree Booster	eta	0.5
	max˙depth	10
	tree˙method	”hist”

Table 3: Hyper parameters for the XGBoost model

C Feature sets

Table 4 presents the list of feature sets tested, along with their corresponding features. The final dataset represents the chosen feature space employed in the analysis.

Table 4: Features in each feature set

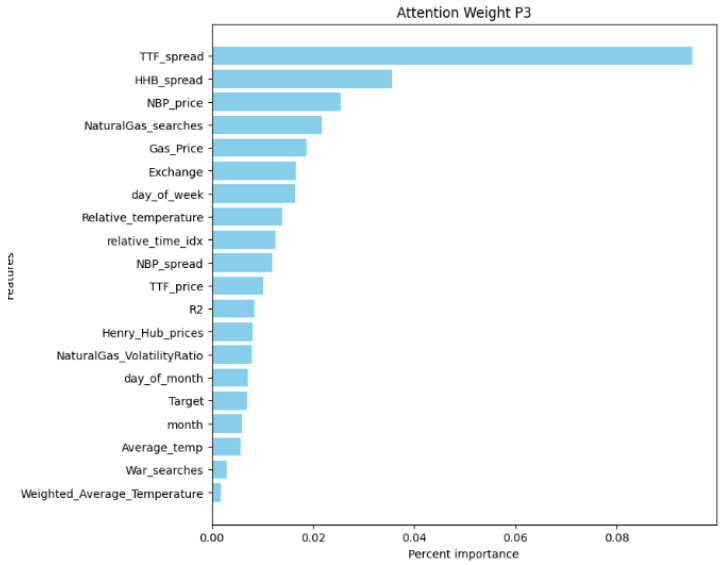
Features	Simple	Base	FS1	FS2	FS3	FS4	FS5	Final
Exchange		✓	✓	✓	✓	✓	✓	✓
EU Clean Spark Spread					✓			✓
Oil Price						✓		✓
Wind Energy					✓			
Average Temp.		✓	✓	✓	✓	✓	✓	✓
Weighted Average Temp.		✓	✓	✓	✓	✓	✓	✓
Fullness Rate				✓				
Net Injection				✓				✓
Shipping Price								✓
NaturalGas'searches			✓					
War searches			✓					✓
TTF Price		✓	✓	✓	✓	✓	✓	✓
TTF Spread		✓	✓	✓	✓	✓	✓	✓
NBP Price		✓	✓	✓	✓	✓	✓	✓
NBP Spread		✓	✓	✓	✓	✓	✓	✓
Henry Hub Price		✓	✓	✓	✓	✓	✓	✓
Henry Hub spread		✓	✓	✓	✓	✓	✓	✓
Target	✓	✓	✓	✓	✓	✓	✓	✓
R2	✓	✓	✓	✓	✓	✓	✓	✓
Relav R2	✓	✓	✓	✓	✓	✓	✓	✓
Oil R2						✓		✓
Relav Temperature		✓	✓	✓	✓	✓	✓	✓
Coal Futures Price						✓		✓
Gazprom							✓	
EOAN							✓	
XOM							✓	✓
EQNR							✓	✓
Gas Price	✓	✓	✓	✓	✓	✓	✓	✓
Sun Power					✓			
CO2 spot					✓			
Day of week	✓	✓	✓	✓	✓	✓	✓	✓
Month	✓	✓	✓	✓	✓	✓	✓	✓
Day of month	✓	✓	✓	✓	✓	✓	✓	✓
Holiday	✓	✓	✓	✓	✓	✓	✓	✓

D Feature importance

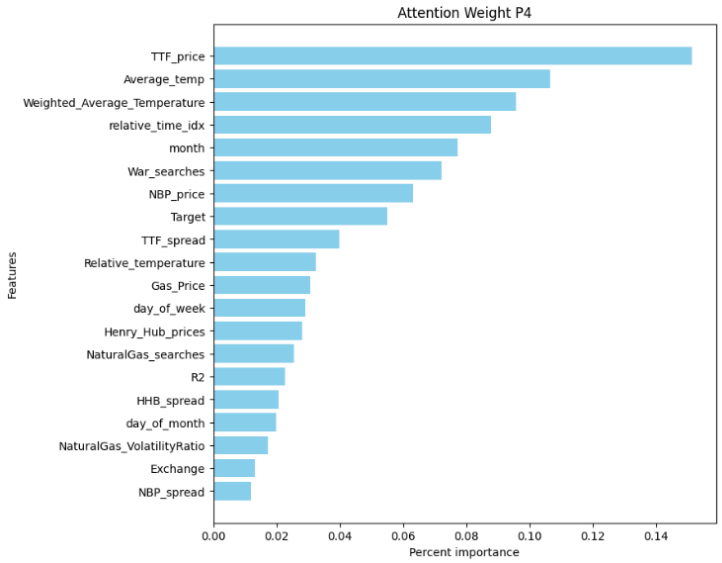
The upcoming section provides an overview of the feature importance of each feature set. This importance is determined by aggregating the attention weights for each forecast within a specific period. For brevity, only the periods discussed in Section 6.1 are showcased to minimize unnecessary space.

D.1 Feature set 1

In Figure 2, the feature importance for period p_3 and p_4 is depicted for feature set 1. This feature set includes additional features "NaturalGas˙searches" and "War˙searches" in addition to the base set.



(a) Period p_3

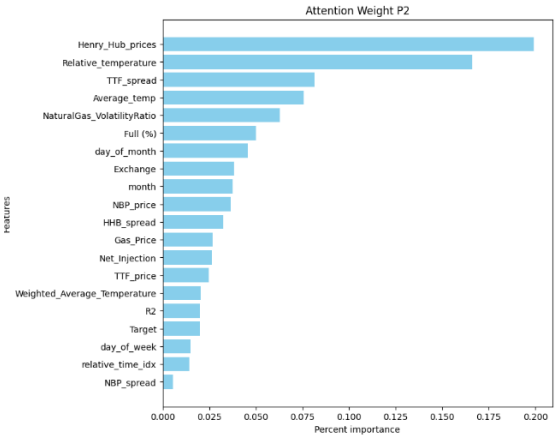


(b) Period p_4

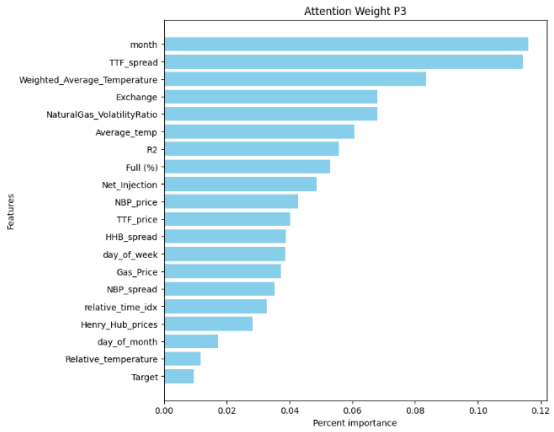
Figure 2: Feature importance for selected periods of Feature set 1

D.2 Feature set 2

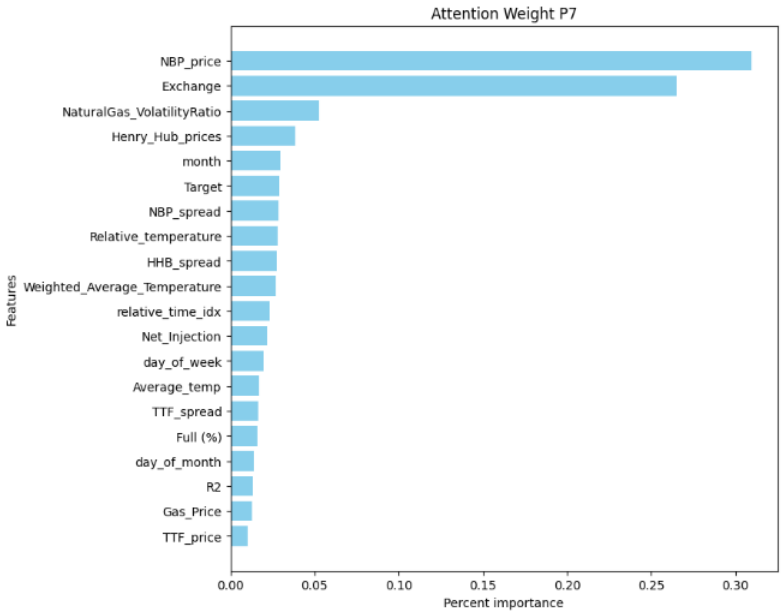
In Figure 3, the feature importance for period p_2, p_3, and p_7 is depicted for feature set 2. This feature set includes additional features "Full(%)" and "Net˙Injection" in addition to the base set.



(a) Period p_2



(b) Period p_3

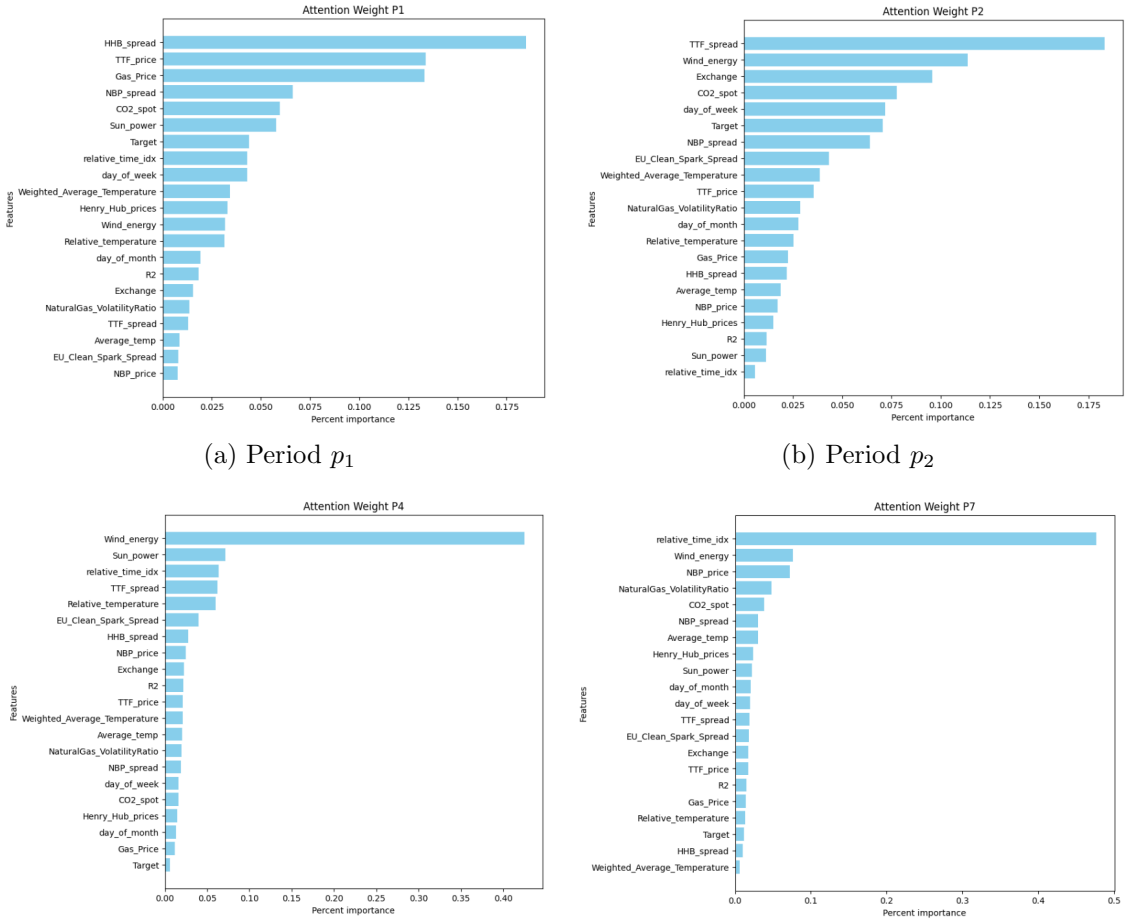


(c) Period p_7

Figure 3: Feature importance for selected periods of Feature set 2

D.3 Feature set 3

In Figure 4, the feature importance for period p_1, p_2, p_4 and p_7 is depicted for feature set 3. This feature set includes additional features "Wind˙energy", "Sun˙power", "EU˙Clean˙Spark˙Spread" and "CO2˙Spot" in addition to the base set.



(a) Period p_1

(b) Period p_2

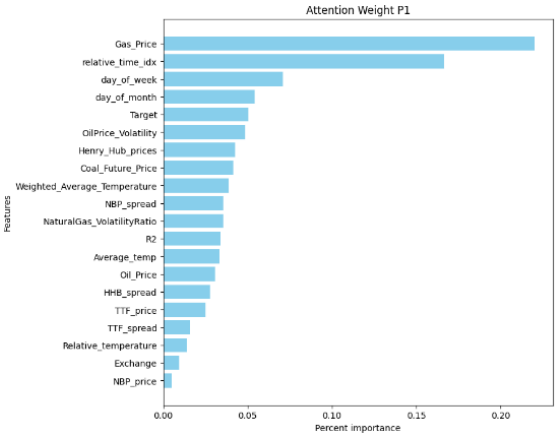
(c) Period p_4

(d) Period p_7

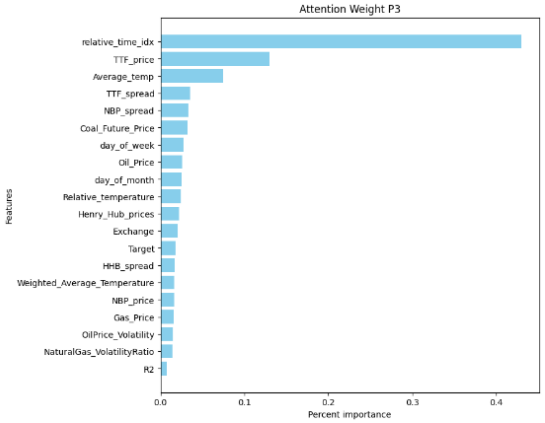
Figure 4: Feature importance for selected periods of Feature set 2

D.4 Feature set 4

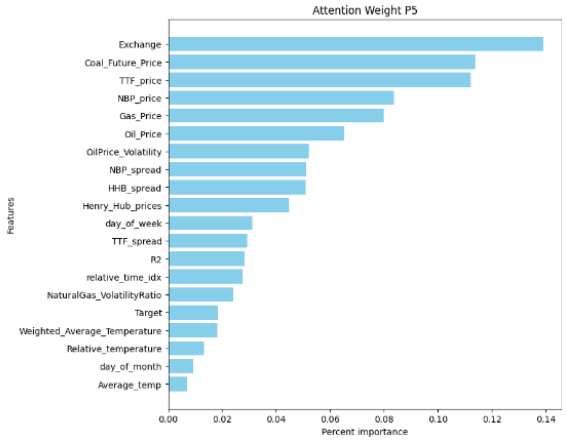
In Figure 5, the feature importance for period p_1, p_3, p_5 and p_8 is depicted for feature set 4. This feature set includes additional features "Coal˙Future˙Price", "Oil˙Price" and "Oil˙Price˙Volatility" in addition to the base set.



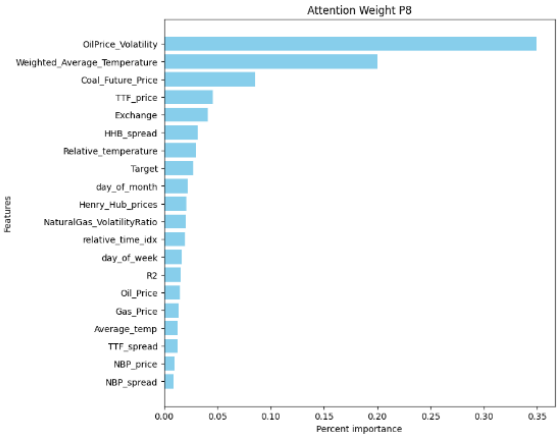
(a) Period p_1



(b) Period p_3



(c) Period p_5

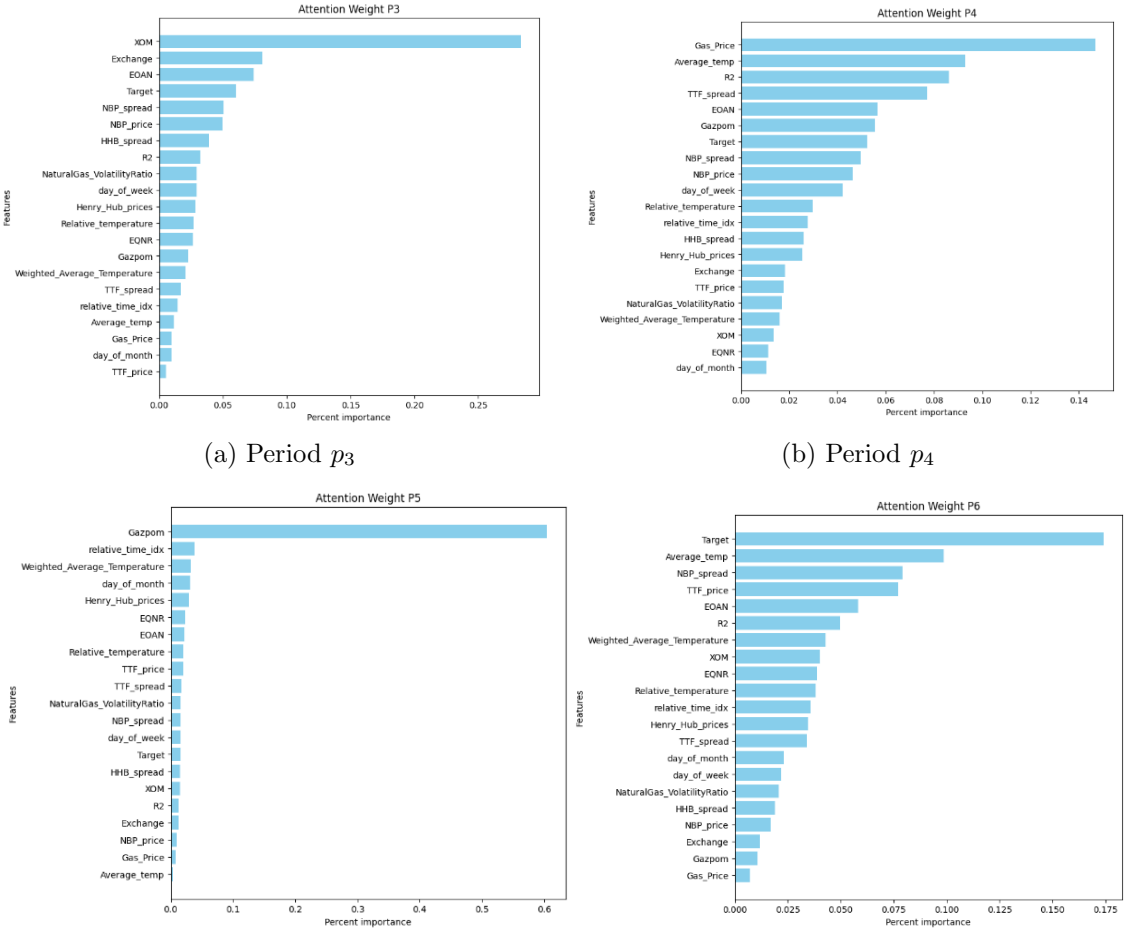


(d) Period p_8

Figure 5: Feature importance for selected periods of Feature set 4

D.5 Feature set 5

In Figure 6, the feature importance for period p_3, p_4, p_5 and p_6 is depicted for feature set 5. This feature set includes additional features "XOM", "EOAN", "EQNR" and "Gazprom" in addition to the base set.



(a) Period p_3

(b) Period p_4

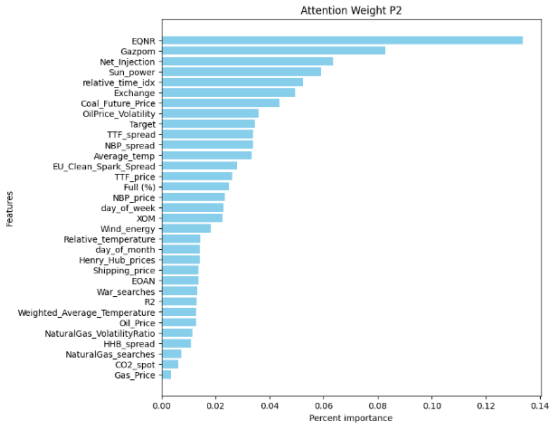
(c) Period p_5

(d) Period p_6

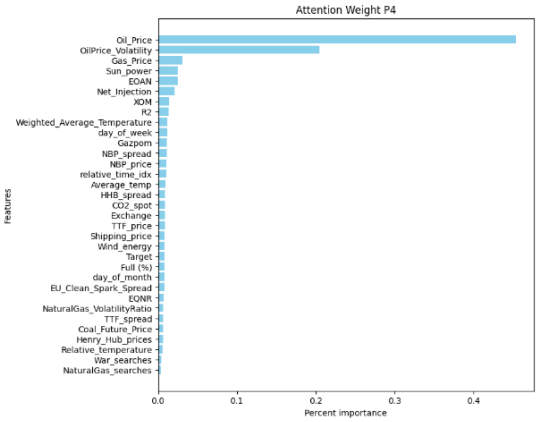
Figure 6: Feature importance for selected periods of Feature set 5

D.6 Full Model

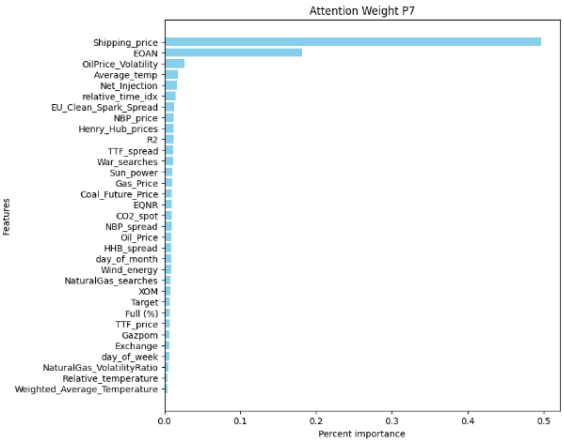
In Figure 7, the feature importance for period p_2, p_4, p_7 and p_8 is depicted for the full model. This feature set includes all features.



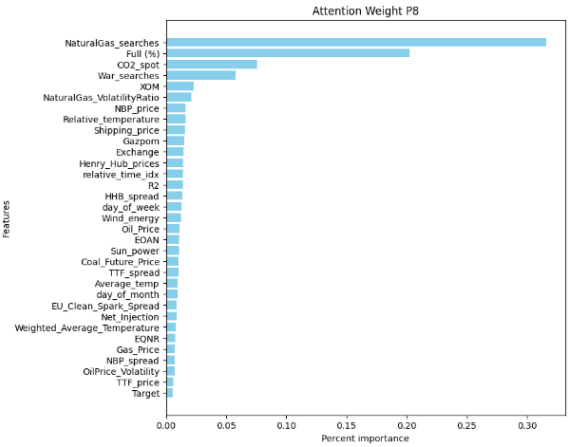
(a) Period p_2



(b) Period p_4



(c) Period p_7



(d) Period p_8

Figure 7: Feature importance for selected periods of the Full model

E Hypotisis test statistics

E.1 Price predictions

Table 5 displays the outcomes of the Diebold-Mariano hypothesis test applied to price predictions. The predictions from TFT are compared with those from ARIMA and XGBoost. Test statistics are highlighted in bold, and corresponding p-values are enclosed in parentheses. Periods with significance below the 5% threshold are marked with **, and periods with significance below the 10% threshold are denoted with *.

Table 5: Diebold Mariano test statistics and p values for the price forecast

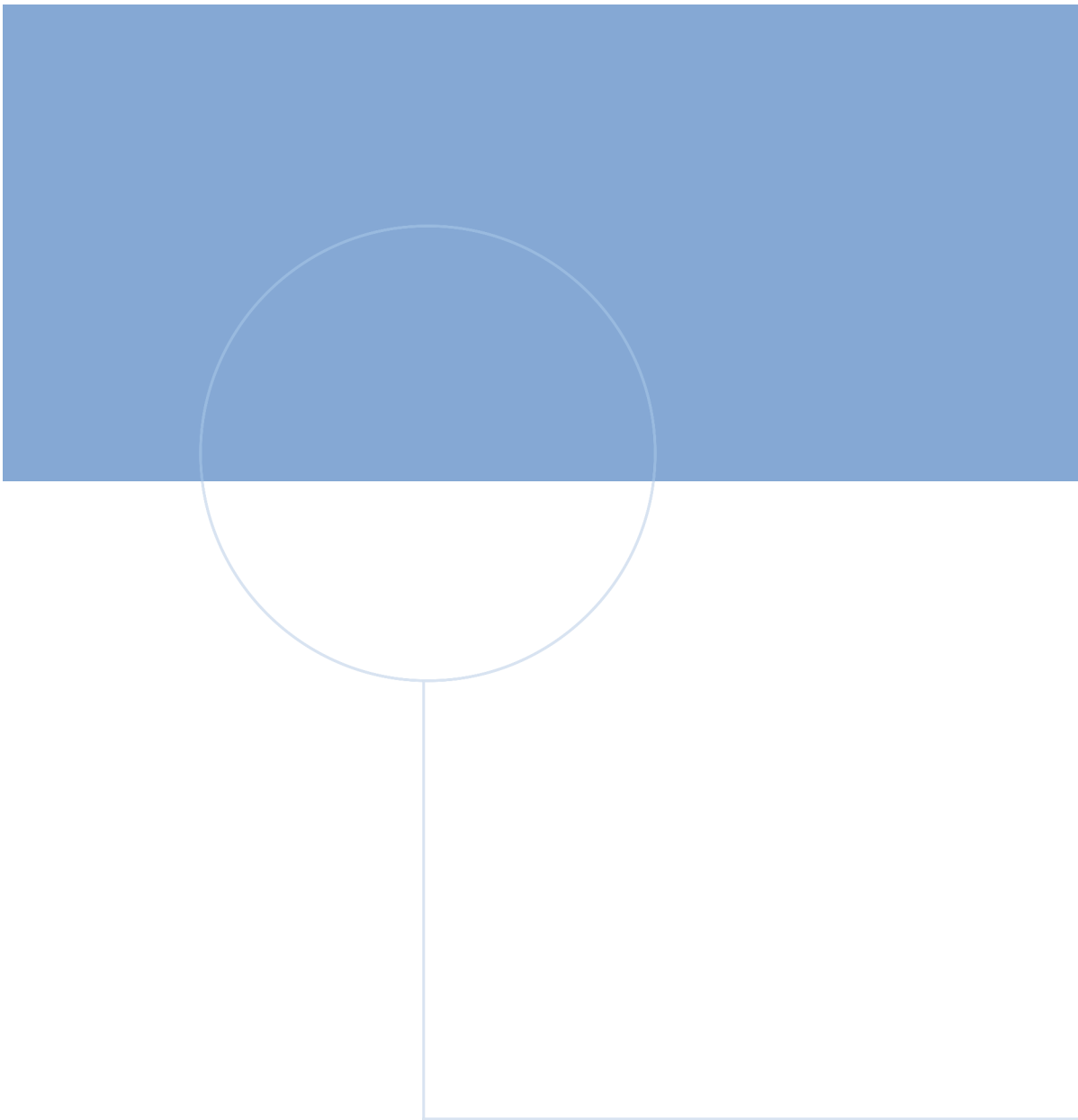
Period	ARIMA	Xgboost
p_1	-2 (0.048)**	-2.4 (0.019)**
p_2	-1.5 (0.13)	-1.38 (0.17)
p_3	-5.6 (6.7×10^{-7})**	-4.7 (0.000014)**
p_4	-1.8 (0.07)*	-1.9 (0.057)*
p_5	-6.91 (3.8×10^{-9})**	-0.41 (0.68)
p_6	-2.1 (0.04)**	-0.69 (0.49)
p_7	-6.0 (1.2×10^{-7})**	-3.2 (0.0023)**
p_8	-1.1 (0.26)	-0.52 (0.61)

E.2 Target predictions

Table 5 displays the outcomes of the Diebold-Mariano hypothesis test applied to return predictions. The predictions from TFT are compared with those from ARIMA and XGBoost. Test statistics are highlighted in bold, and corresponding p-values are enclosed in parentheses. Periods with significance below the 5% threshold are marked with **, and periods with significance below the 10% threshold are denoted with *.

Table 6: Diebold Mariano test statistics and p values for the return forecast

Period	ARIMA	XGBoost
p_1	-2.27 (0.027)**	1.23 (0.22)
p_2	-2.15 (0.036)**	-0.46 (0.65)
p_3	-3.8 (0.00031)**	-2.42 (0.019)**
p_4	-2.3 (0.026)**	-0.61 (0.55)
p_5	-0.61 (0.54)	1.56 (0.13)
p_6	-2.79 (0.007)**	0.04 (0.97)
p_7	-4.6 (0.000027)**	-1.85 (0.07)*
p_8	-0.66 (0.51)	-0.55 (0.58)



NTNU

Norwegian University of
Science and Technology