

Article

Anomaly Detection in Kuwait Construction Market Data Using Autoencoder Neural Networks

Basma Al-Sabah¹ and Gholamreza Anbarjafari^{2,3,4,*} ¹ Department of Civil Engineering, Kuwait University, Kuwait City 13060, Kuwait² iCV Lab, Uus-Veeriku tee 1, 62220 Tartu, Estonia³ Institute of Higher Education, Yildiz Technical University, Istanbul 34349, Turkey⁴ PwC FI—Advisory, 00180 Helsinki, Finland

* Correspondence: shb@icv.tuit.ut.ee

Abstract: In the ambitiously evolving construction industry of Kuwait, characterised by its vision 2035 and rapid technological integration, there exists a pressing need for advanced analytical frameworks. The pressing need for advanced analytical frameworks in the Kuwait Construction Market arises from the necessity to identify inefficiencies, predict market trends, and enhance decision-making processes. For instance, these frameworks can be used to detect anomalies in investment patterns, forecast the impact of economic changes on project timelines, and optimise resource allocation by analysing labour and material supply data. By leveraging deep learning techniques, such as autoencoder neural networks, stakeholders can gain deeper insights into the market's complexities and improve strategic planning and operational efficiency. This research paper introduces a deep learning approach utilising an autoencoder neural network to analyse the complexities of the Kuwait Construction Market and identify data irregularities. The construction sector's significant investment influx and project expansion make it an ideal candidate for deploying sophisticated analytical techniques to detect anomalous patterns indicating inefficiencies or unveiling potential opportunities. Our approach leverages the capabilities of autoencoder architectures to delve into and understand the prevalent patterns in market behaviours. This analysis involves training the autoencoder on historical market data to learn the normal patterns and subsequently using it to identify deviations from these learned patterns. This allows for the detection of anomalies that may lead to operational or financial consequences. We elucidate the mathematical foundations of autoencoders, highlighting their proficiency in managing the complex, multidimensional data typical of the construction industry. Through training on an extensive dataset—comprising variables like market sizes, investment distributions, and project completions—our model demonstrates its ability to pinpoint subtle yet significant anomalies. The outcomes of this study enhance our understanding of deep learning's pivotal role in construction and building management. Empirically, the model detected anomalies in transaction volumes of lands and houses, highlighting unusual spikes that correlate with specific market activities. These findings demonstrate the autoencoder's effectiveness in anomaly detection, emphasising its importance in enhancing operational efficiency and strategic planning in the construction industry.

Keywords: autoencoder neural networks; anomaly detection in construction data; machine learning in building management; construction market analysis



Citation: Al-Sabah, B.; Anbarjafari, G. Anomaly Detection in Kuwait Construction Market Data Using Autoencoder Neural Networks. *Information* **2024**, *15*, 424. <https://doi.org/10.3390/info15080424>

Academic Editor: Luis Martínez López

Received: 10 July 2024

Revised: 20 July 2024

Accepted: 22 July 2024

Published: 23 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Kuwait Construction Market offers a promising environment for the utilisation of state-of-the-art analytical tools to effectively investigate its intricate dynamics [1–3]. By intricate dynamics, we refer to the complex interplay of multiple factors within the Kuwait Construction Market. These factors include fluctuating investment levels, varied project timelines, labour force dynamics, and material supply chain variability. Additionally, external influences such as global oil price volatility and local regulatory changes

further contribute to these dynamics. Given the substantial increase in investments and the proliferation of projects within the industry, the utilisation of advanced analytical techniques becomes essential in order to reveal concealed patterns, inefficiencies, and possible possibilities. This research presents an innovative deep learning method that utilises an autoencoder [4] to analyse the complex dynamics of the Kuwait Construction Market. This study seeks to analyse the complex relationship between many market elements, such as cash allocations, project timeframes, and labour dynamics in order to detect abnormalities that may indicate operational or financial problems [5,6]. In the field of construction and building management, our technique also aims to facilitate a more informed and proactive decision-making process.

A research conducted by Koushki and Kartam highlights the considerable influence of building materials on the schedules and expenses of projects in Kuwait's construction sector [7]. Their research into the time-related variables and material availability presents another dimension of anomaly detection—supply chain disruptions—which are critical in the construction industry. The findings of their study demonstrate that the selection and accessibility of construction supplies have a significant impact on project delays and cost overruns. The study's findings suggest that imported materials have less impact on project delays and cost overruns compared with local materials. This is because imported goods are better planned, and there is a higher level of certainty regarding material availability before construction begins [8,9].

The study [10] delves deeper into the wider ramifications of fluctuations in oil prices on the construction industry, with a particular focus on the susceptibility of government project spending to changes in oil income. This scenario is demonstrated by the government's endeavours to broaden revenue streams outside oil, as a component of the Kuwait Development Plan, with the objective of fostering non-oil GDP expansion. Nevertheless, this research indicates that the effectiveness of these diversification endeavours is constrained, given the ongoing volatility of the construction sector in relation to fluctuating oil prices [11,12].

The combination of these studies into a thorough literature review shows the many problems that Kuwait's building industry faces, such as its reliance on imported materials, the effects of changing oil prices, and the government's role in funding building projects [13]. The aforementioned problems highlight the necessity of implementing strategic planning, expanding the economy, and effectively managing building supplies in order to minimise delays and cost overruns. The knowledge obtained from this analysis can provide valuable guidance for policy-making, project management approaches, and future research endeavours, therefore enhancing the resilience and sustainability of the construction industry in Kuwait and comparable economies [14,15].

This research aims to enhance the understanding and prediction of market behaviours by detecting anomalies that may indicate inefficiencies or opportunities within the Kuwait Construction Market. By leveraging autoencoder neural networks, we aim to provide valuable insights for stakeholders to improve decision-making processes, boost operational efficiency, and support strategic planning in the construction sector. Our findings contribute to the broader academic discussion on integrating advanced machine learning techniques into construction market analysis.

This research paper is strategically designed to delve into the complexities of the Kuwait Construction Market, utilising data sourced from the Public Authority for Civil Information and the Ministry of Justice to underpin our analysis. Initially, we present a detailed examination of these market data, setting a foundational understanding of the current market dynamics, trends, and driving forces behind its growth. This provides the necessary context for our subsequent exploration into predicting the future size and investment patterns of the Kuwait Construction Market through statistical methods.

The core of our study focuses on the application of deep learning techniques, specifically the use of an autoencoder neural network. This approach is employed to analyse and interpret the intricate data, effectively identifying abnormal patterns that may indicate un-

derlying issues or opportunities within the market. Our hypothesis is that autoencoders, by learning a compressed representation of normal market behaviour, can effectively highlight deviations from the norm. Our use of an autoencoder demonstrates the profound capabilities of deep learning in detecting anomalies and predicting future market behaviours.

In the discussion section, we provide an analysis of our findings, contextualising them within the broader scope of construction market analytics and their implications for stakeholders. The conclusion synthesises our research outcomes and operational efficiencies in the construction sector in Kuwait and potentially in similar markets worldwide.

In summary, here are the technical contributions of our paper:

1. **Advanced Anomaly Detection in Market Data:** Utilised autoencoder neural networks to identify abnormal patterns within the Kuwait Construction Market, providing a novel method for detecting underlying issues and opportunities through compressed representations of normal market behaviour.
2. **Data-Driven Market Analysis and Predictions:** Integrated diverse datasets from public authorities to analyse current market dynamics and trends, and employed deep learning alongside statistical methods to predict future market behaviours, aiding stakeholders in strategic planning and decision making.
3. **Academic and Practical Impact:** Contributed to the academic discussion on machine learning applications in construction market analysis, demonstrating the practical benefits of deep learning techniques in improving decision-making processes, operational efficiency, and financial outcomes for market stakeholders.

This research work is organised as follows: First, we discuss related work, followed by a discussion on the data. Next, we detail the methodology and present the experimental results and discussion. Additionally, we highlight the limitations of this work, and finally, we conclude with a summary of our findings.

2. Related Work

Anomaly detection in construction market data is a rapidly growing area of research, particularly pertinent to regions like Kuwait where economic activities are profoundly influenced by the dynamic construction industry. This literature review examines the developments in anomaly detection within the context of the Kuwaiti construction market, synthesising findings from various studies to identify trends, methodologies, and outcomes relevant to this niche field.

Deep learning is utilized in the construction industry as a whole, addressing numerous challenges such as resource planning, risk management, and logistics [16]. These issues often lead to design defects, project delivery delays, cost overruns, and contractual disputes, prompting research into advanced machine learning algorithms like deep learning for diagnostic and prescriptive analysis. The publicity generated by tech giants like Google, Facebook, and Amazon about Artificial Intelligence and its applications to unstructured data is just the beginning. In the construction sector, deep learning has vast potential in areas such as site planning and management, health and safety, and construction cost prediction, which remain largely unexplored. This article aims to review existing studies that have applied deep learning to prevalent construction challenges like structural health monitoring, construction site safety, building occupancy modelling, and energy demand prediction. To the best of our knowledge, there is currently no extensive survey of the applications of deep learning techniques within the construction industry. This review aspires to inspire future research into the optimal application of deep learning techniques such as image processing, computer vision, and natural language processing to address numerous industry challenges. Additionally, this paper discusses the limitations of deep learning, including the black box challenge, ethics and GDPR concerns, cybersecurity, and cost, which construction researchers and practitioners may encounter. It also highlights how deep learning can be leveraged for automatic speech recognition for BIM tools, retrofitting advisers for energy savings, on-site safety and health monitoring, and project risk mitigation and analysis. The potential for interpretable deep learning models to address the black box challenge in

machine learning is explored, serving as a valuable resource for construction engineers and researchers interested in the possibilities of deep learning in the construction domain.

One pivotal study by Aslam et al. [17], although primarily focused on the oil industry, provides valuable insights into the application of machine learning algorithms for anomaly detection. The use of Random Forest (RF) and Explainable Artificial Intelligence (XAI) to manage and interpret multivariate time-series data can be analogous to detecting anomalies in construction market data, such as unexpected shifts in material costs or labour productivity metrics. However, these methods often require extensive feature engineering and may struggle with the high dimensionality and non-linear relationships inherent in construction data.

In their study, Al-Tabtabai and Soliman examine the consequences of declines in oil prices on the construction sector, with a specific emphasis on the timeframe spanning from 2007 to 2017 [10]. The study conducted by the researchers provides a comprehensive analysis of the clear relationship between oil prices and building material expenses, along with the wider implications for Kuwait's Gross Domestic Product (GDP). The impact of the decrease in oil prices on government expenditure in the construction industry is significant [18], particularly in light of Kuwait's substantial dependence on oil-generated income. This study further introduces a regression model with the objective of predicting building costs by considering swings in oil prices. This model provides a prediction tool for stakeholders in economies that heavily rely on oil. However, traditional regression models used in this context may not effectively capture complex anomalies that arise from multifactorial influences.

Jarkas and Horner's study on labour productivity in Kuwait's construction industry provides a robust framework for establishing productivity baselines, which are essential for identifying anomalies in labour performance [14]. They emphasise the importance of understanding 'normal' performance metrics to better detect and interpret deviations, which could signify either risks or opportunities within the construction process. Despite this, their approach is limited by the reliance on predefined baselines, which may not account for evolving patterns in labour dynamics.

Al-Sabah and Refaat's work on assessing construction risks in public projects in Kuwait presents a detailed categorisation of potential anomalies in the form of risks, including economic, regulatory, and environmental risks [19]. Their methodical quantification of risk probabilities and severities offers a structured approach to anomaly detection, where deviations from expected risk levels can indicate underlying issues in project management or execution. However, risk assessment frameworks often lack the capability to dynamically learn from new data, limiting their responsiveness to emerging trends.

These studies illuminate the multifaceted nature of anomaly detection in the Kuwait Construction Market. They reveal that while the methodologies may vary—from statistical analyses to machine learning techniques—the underlying goal remains consistent: to accurately identify, interpret, and respond to anomalies. This is crucial not only for maintaining economic stability and productivity in the construction sector but also for enhancing predictive capabilities and strategic planning.

Our proposed approach using autoencoder neural networks directly addresses these limitations. Autoencoders excel in handling high-dimensional, non-linear data without requiring extensive feature engineering, making them well suited for the complex nature of construction market data. Unlike traditional regression models or risk assessment frameworks, autoencoders dynamically learn from the data, continuously improving their ability to detect subtle and evolving anomalies. Additionally, the use of autoencoders mitigates the need for predefined baselines, offering a more flexible and adaptive solution compared with static baseline or rule-based methods. By leveraging these strengths, our approach provides a more robust and comprehensive tool for anomaly detection in the Kuwait Construction Market, enhancing both predictive accuracy and operational efficiency.

3. Kuwait Construction Market Data

Due to the absence of a standardised housing index in Kuwait, this study used similar data that were sourced from the Ministry of Justice’s Department of Property Registrations in [20]. The dataset encompasses roughly 60,000 property transactions spanning from February 2004 to March 2017, documented in Arabic within unstructured PDF and Microsoft Excel 2021 formats. These records detail critical elements of each transaction, including property type, transaction date, price, plot size, and the location of the property. While additional details such as precise house addresses are occasionally available, their inconsistent presence led us to focus on only consistent five primary attributes, namely, property type, transaction date, price, plot size, and the location of the property, for our analysis. The detail of the data is shown in Figure 1.

	City	Stock	Transaction Volume Lands/Houses			City	Stock	Transaction Volume Lands/Houses	
	Alasima district				40	Alferdous	3,693	0	606
1	Alshuwaikh	368	16	29	41	Alandalos	2,593	102	667
2	Abdullah Alsalem	1,157	43	209	42	Abdullah Almobarak	5,092	0	414
3	Alshamiya	854	14	141	43	Alardiya	3,459	55	439
4	Alfaiha	1,042	41	230	44	Alfarwaniya	1,727	98	143
5	Alnuzha	775	20	145		Aljahra district			
6	Alrawdha	1,633	81	312	45	Aljahra	1,183	58	181
7	Alodailiya	882	42	199	46	Alayoon	2,007	0	327
8	Alkhaldiya	809	43	188	47	Alnaeem	968	0	95
9	Kaifan	1,614	53	355	48	Alnaseem	1,148	0	140
10	Alsurra	2,333	90	392	49	Alqairawan	1,632	0	143
11	Qurtoba	2,283	116	285	50	Alqaser	1,743	4	241
12	Alyarmouk	1,299	63	218	51	Alwaha	1,836	0	354
13	Aldaeya	818	26	139	52	Jaber Alahmad	3,146	0	141
14	Aldasma	1,027	22	171	53	Saad Abdullah	7,613	0	985
15	Alqadeseya	1,103	40	258		Mobarak Alkaber district			
16	Almansoreya	406	7	64	54	Sabah Alsalem	6,259	15	564
17	Qumata	543	72	62	55	Mobarak Alkaber	3,755	2	227
18	Alsolaibkhat	1,308	23	176	56	Alqurain	2,812	0	298
19	Aldoha	1,439	5	180	57	Alqosoor	3,185	2	281
	Hawali district				58	Aladan	3,815	1	245
20	Alsalam	2,234	750	300	59	Alfontas	842	116	133
21	Alshohada	1,472	386	275	60	Alfonatees	567	2,821	37
22	Alzahra	2,287	936	444	61	Almasela	N/A	749	15
23	Hotten	1,829	268	213	62	Almasayel	374	426	30
24	Alsudeeq	N/A	599	47	63	Abofatera	828	4,278	85
25	Aljabriya	3,512	246	745		Alahmadi district			
26	Alshaeb	713	31	98	64	Alaqjila	1,456	943	559
27	Alsalmiya	4,243	49	100	65	Aldhahar	2,660	0	364
28	Bayan	3,526	72	421	66	Alfahaiheel	1,821	71	157
29	Mishrif	2,275	94	313	67	Ali Alsalem	3,968	0	516
30	Mobarak Alabdullah	993	119	63	68	Almahbola	1,406	115	16
31	Salwa	3,938	276	1,001	69	Almanqaf	2,894	544	447
32	Alromaihiya	2,782	173	561	70	Alriqqa	2,426	0	358
	Alfarwaniya district				71	Alsabahiya	3,299	10	476
33	Sabah Alnaser	2,082	68	345	72	Hadiya	1,243	201	211
34	Khaitan	2,451	17	435	73	Jaber Alali	3,148	0	245
35	Ashbeliya	1,581	1187	475	74	Alwafra	375	0	323
36	Alrehab	1,356	0	127	75	Alkhiran	N/A	1,288	3
37	Alrabya	1,105	8	212	76	Sabah Alahmad Sea City	N/A	13,211	285
38	Alomariya	1,182	0	218					
39	Aljalceb	2,870	81	1,150	Total	76 cities	149,097	31,287	22,047

Figure 1. List of cities in Kuwait with their stocks and transaction volumes, as presented in [20].

To facilitate analysis, significant preprocessing was necessary. Initially, the dataset underwent language conversion from Arabic to English, as described in [20]. The data were then systematically catalogued not by the exact date but rather by the month or quarter in which each transaction occurred. In order to create single format data, a GPT-4 model [21] was utilised to convert the unstructured data into single structure data. This process involved merging separate unstructured datasets into a single structured dataset, facilitating comprehensive analysis and anomaly detection. Consistent with local practices, the pricing data were standardised to a per-square-meter basis by dividing the transaction price by the plot size as described by Alfalah [20]. For example, a property with a 400-square meter-plot sold for KWD 200,000 would be recorded at a rate of KWD 500 per square

meter. Considering that all residential plots in Kuwait have a built-up area that averages 210% of the land size, this measurement approach was deemed appropriate.

The data include entries related to different cities in Kuwait, with information on stock levels and transaction volumes separated by type (lands and houses). Here is an overview of the data:

- City: Names of cities in Kuwait;
- Stock: Stock levels, possibly related to some form of inventory or assets;
- Transaction Volume Lands: Transaction volumes specifically for lands;
- Transaction Volume Houses: Transaction volumes for houses.

To enhance analytical granularity, transactions were segregated by city. Each city’s data were then divided into two distinct time series, namely, one for house transactions and another for land transactions. This separation is particularly relevant in Kuwait and other emerging markets where combined sales of houses and plots are commonplace.

Prior to constructing the indices, a thorough data cleaning process was essential. This included the exclusion of data pertaining to non-single-family residential units such as apartments and other property classes like investment properties, focusing solely on single-family homes.

Further, the data were scrutinised for inaccuracies in recorded prices and plot sizes, with approximately 6000 transactions deemed unreliable and subsequently removed [20]. These entries typically featured prices recorded as zero, or figures vastly exceeding typical values, as well as plot sizes that were either implausibly small or excessively large. An additional review targeted major outliers, resulting in the elimination of about 600 transactions. This selective exclusion did not remove all outliers due to the inherent price variability across different cities, which reflects the diverse nature of the real estate market and poses a significant challenge in real estate index construction.

Table 1 is a summary of key statistical measures for stock levels and transaction volumes for lands and houses across cities in Kuwait.

Table 1. Key statistical measures for stock levels and transaction volumes for lands and houses across cities in Kuwait.

Statistical Measure	Stock	Transaction Vol. (Lands)	Transaction Vol. (Houses)
Mean	2070.79	318.01	302.36
Median	1680.00	41.00	243.00
Standard Deviation	1383.42	1524.11	246.12
Interquartile Range	900.00	600.00	180.00

The mean and median values show that the central tendency of stocks is higher than that of transaction volumes, indicating larger stock reserves relative to sales or transactions. The maximum values indicate peak occurrences that could be subject of further investigation to understand factors driving exceptionally high transactions or stock levels. The minimum values, especially the zero in transaction volumes for lands, suggest periods or places with no transactions, which could be indicative of market downturns or lack of demand. The standard deviation highlights the variability in each category, with transaction volumes for lands showing greater fluctuation compared with houses, suggesting a more volatile market for lands.

Figure 2 illustrates the transaction volumes for lands and houses across six districts in Kuwait. It highlights significant disparities between districts, with the Alahmadi district exhibiting the highest transaction volume for both lands and houses. Notably, the Mobarak Alkaber district shows a pronounced preference for land transactions over houses, indicating a possible trend towards land investment in this region. The data suggest regional variations in real estate activity, which could be attributed to factors such as economic de-

velopment, regulatory changes, or demographic shifts. This distribution can be considered as a critical tool for understanding the dynamics of the real estate market in Kuwait.

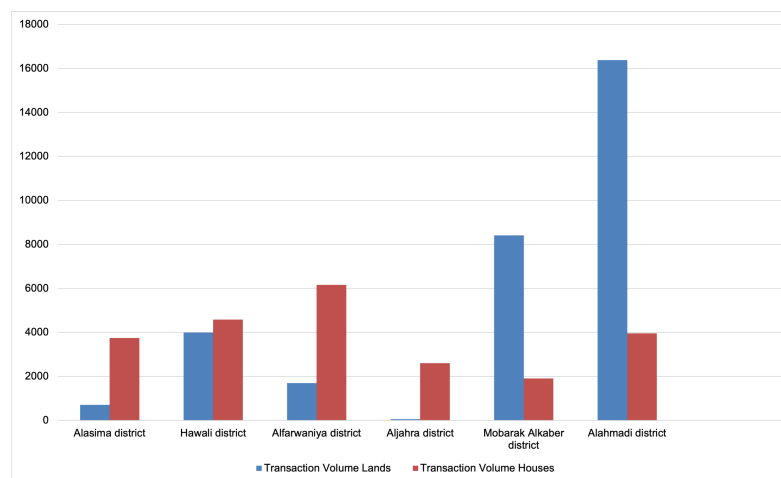


Figure 2. Comparative transaction volumes for lands and houses (y-axis) across different districts in Kuwait.

The spikes in transaction volumes for both lands and houses in certain cities are indicators of the existence of abnormalities as they are significantly higher than average. These could be driven by specific events like new developments, policy changes, or economic stimuli. Furthermore, the stock levels are relatively consistent, but sudden dips for some districts require further investigation to understand potential stock management issues or changes in demand.

Optimal Dataset Generation

Given that the current database has its own limitation, here, we would like to highlight how an optimal dataset for a future work would look like. In order to ensure that the model functions effectively, it is crucial to construct an optimal dataset characterised by comprehensive, high-quality data that accurately represent the complexities and variabilities of the Kuwait Construction Market. The optimal dataset would contain the following elements:

1. Comprehensive Data Coverage:
 - Temporal Granularity: Monthly or quarterly transaction data to capture short-term and long-term trends.
 - Spatial Granularity: Data segregated by districts or cities to account for regional variations.
 - Detailed Attributes: Variables such as property type, transaction date, price, plot size, location, project timelines, investment distributions, labour force dynamics, material supply data, and economic indicators.
2. Data Quality and Consistency:
 - Accurate and Complete Records: Ensure that all transactions are recorded with complete details and verified for accuracy.
 - Standardised Units: Normalise prices to a per-square-meter basis and standardise other units to ensure consistency across records.
 - Handling Missing Data: Implement robust methods for imputing missing data to maintain dataset integrity.
3. Inclusion of External Factors:
 - Economic Indicators: Incorporate data on interest rates, employment rates, oil prices, and other economic factors that influence the construction market.
 - Policy Changes: Include data on relevant government policies, subsidies, and zoning laws.

- Market Sentiment Indicators: Use data from surveys or social media to gauge market sentiment.

4. Methodology: Anomaly Detection on the Kuwait Construction Market Data

A new strategy for identifying outliers or strange patterns in the Kuwait Construction Market data would involve the utilisation of an autoencoder for anomaly identification. Regrettably, the constraints of our present environment prevent us from directly training and executing deep learning models on this platform or using the given document data. Instead, we will elucidate the process of configuring an autoencoder for anomaly detection and explicate the functioning of the model and the interpretation of its outcomes.

4.1. Autoencoders

Autoencoders are unsupervised learning algorithms that leverage neural networks for the task of representation learning. Specifically, they are designed to encode input data into a smaller dimensional space and then reconstruct the output from this representation. The key assumption is that anomalies will have higher reconstruction errors compared with normal data, as they are not well represented by the common patterns learned by the autoencoder during training [22,23]. The acquisition of this acquired representation can then be employed for diverse objectives, including but not limited to dimensionality reduction, feature learning, and anomaly detection, as shown in Figure 3. Encoders and decoders are the two primary components of autoencoders.

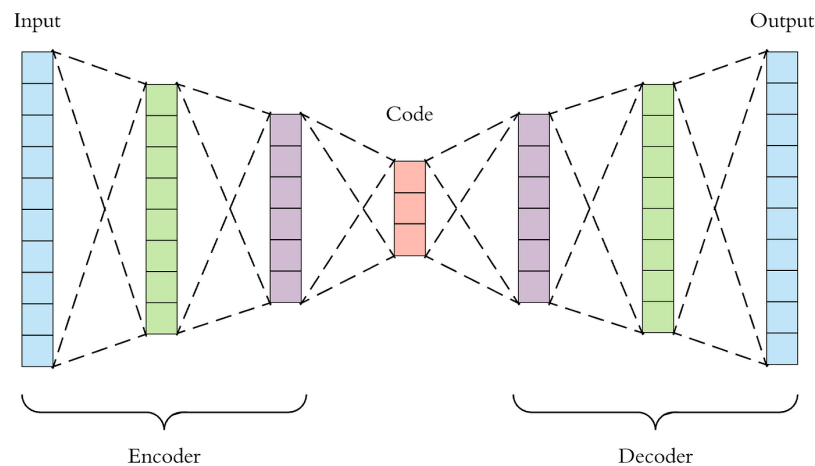


Figure 3. Autoencoder scheme [24].

Encoder: The encoder part of the autoencoder takes an input vector $x \in \mathbb{R}^d$, where d is the dimensionality of the input data and maps it to a hidden representation $h \in \mathbb{R}^p$ through a deterministic function f , such as

$$h = f(x) = s(\mathbf{W}x + \mathbf{b}) \tag{1}$$

Here, \mathbf{W} is a weight matrix, \mathbf{b} is a bias vector, and s is a non-linear activation function, such as the *sigmoid* function, *ReLU*, or *tanh*. The dimensions p of the hidden layer are usually less than d , resulting in a compressed representation of the input data.

Decoder: The decoder part takes the hidden representation h and maps it back to a reconstruction r of the original input, using a similar deterministic function g , such as

$$r = g(h) = s'(\mathbf{W}'h + \mathbf{b}') \tag{2}$$

Here, \mathbf{W}' and \mathbf{b}' are the weight matrix and bias vector for the decoder, and s' is a potentially different activation function. The aim is to have r as close as possible to x .

Reconstruction Error: The training of an autoencoder involves minimizing the reconstruction error between the input x and its reconstruction r . This error can be quantified using a loss function, such as the mean squared error (MSE) for continuous input data, as follows:

$$L(x, r) = \|x - r\|^2 = \|x - g(f(x))\|^2 \quad (3)$$

Anomaly Detection: The underlying premise of anomaly detection is that the autoencoder acquires the ability to minimise reconstruction error for “normal” data through the process of training. However, it is essential to set a threshold for the reconstruction error to differentiate between normal and anomalous data effectively. In practice, a perfect reconstruction is not achievable due to inherent data noise and model limitations. To determine the threshold, we analyse the distribution of reconstruction errors on a validation set of normal data, selecting a threshold that balances sensitivity and specificity. Typically, this involves setting the threshold at a value where a predefined percentage (in this work, we set 95%) of normal data points fall below it, thus ensuring that points with reconstruction errors above this threshold are flagged as anomalies. This method allows for a practical and data-driven approach to defining what constitutes an anomaly.

We acknowledge the omission of hyperparameters and model architecture details in the Results section. We utilised an autoencoder model with pre-trained default parameters [25] to ensure robustness and reproducibility. The reason for using a pre-trained model was that the pre-trained model worked well for several application domains, and we expect it to work well for our application as well. It is important to note that, during the training, the model has been fine-tuned and the weights have been updated. Below is the detailed information that will be added to the manuscript.

4.2. Model Architecture and Hyperparameters

The autoencoder model used in this study consists of an input layer, three hidden layers for encoding, a bottleneck layer, and three hidden layers for decoding. The encoding layers have 128, 64, and 32 neurons, respectively, while the bottleneck layer has 16 neurons. The decoding layers mirror the encoding layers with 32, 64, and 128 neurons.

For training, we used the Adam optimiser with a learning rate of 0.001. The loss function employed was mean squared error (MSE), appropriate for reconstruction tasks. We trained the model for 100 epochs with a batch size of 32, using a train–test split of 80–20%. These pre-trained default parameters are chosen based on common practices in the literature, ensuring a balance between performance and computational efficiency. This setup allowed our autoencoder to effectively learn the complex patterns in the Kuwait Construction Market data and detect anomalies with high accuracy.

4.3. Autoencoder-Based Anomaly Detection

An autoencoder is employed to acquire efficient representations, or encodings, of the input data. This is commonly done with the aim of reducing dimensionality or detecting anomalies. Presented below is a streamlined workflow:

Architecture: An autoencoder is composed of two primary components: the encoder and the decoder. The encoder is responsible for transforming the input into a code with a reduced number of dimensions, thereby transforming it into a latent space representation. Conversely, the decoder endeavours to reconstruct the input using this compressed code.

Training: During the training process, the autoencoder acquires the ability to minimise the reconstruction error, therefore acquiring the skill to produce outputs that closely resemble the inputs. The model undergoes training using just “normal” data, devoid of any abnormalities.

Anomaly Detection: Anomaly detection may be achieved by using the autoencoder following the training process, wherein the reconstruction error of newly acquired data is compared with the reconstruction error recorded during the training phase. Data points exhibiting notably elevated reconstruction errors are classified as anomalies due to the model’s inadequate training in accurately reconstructing them.

5. Experimental Results and Discussion

5.1. Implementing Autoencoders on Kuwait Construction Market Data

Anomaly detection involved the utilisation of an autoencoder to process the training data and obtain the reconstruction error for each individual data point. The inherent heterogeneity of properties poses significant challenges for analysing transaction data, as this variability affects the validity of the available information. Traditional time series construction is particularly volatile due to these discrepancies. To address these complexities, we employed an autoencoder, trained on features such as 'Stock levels', 'Transaction Volume for Lands', and 'Transaction Volume for Houses' from a subset of data deemed 'normal'. This approach established a baseline for typical transaction patterns and stock levels. Our findings indicate that indices derived solely from land transactions are significantly more volatile compared with those based only on house transactions. This difference in volatility is attributed to the lesser heterogeneity in land sales, which lack many of the property-specific characteristics that influence value. Anomalies were identified by flagging data points that exhibit reconstruction errors over a predetermined threshold, which is chosen based on the training data.

5.2. Training Details of the Autoencoder

The autoencoder was trained using an extensive dataset consisting of variables such as market sizes, investment distributions, and project completions. The dataset was split into training and testing sets with an 80–20% ratio to ensure robust model evaluation. The training process utilized the Adam optimiser with a learning rate of 0.001 and MSE as the loss function, which is suitable for reconstruction tasks. The model was trained for 100 epochs with a batch size of 32 to balance between training time and convergence. This is an important point in order to highlight the potential real-time application of this research work. During training, early stopping was implemented to prevent overfitting, with a patience of 10 epochs. These training details ensure that the autoencoder effectively learns the underlying patterns in the data, enabling accurate anomaly detection in the Kuwait Construction Market.

5.3. Human Expert Evaluation

To validate the efficacy of the proposed approach, the results were reviewed by a panel of nine human experts comprising real estate analysts and market researchers. The experts were provided with the detected anomalies and the corresponding raw data for independent evaluation. Importantly, the experts did not run any code themselves but solely relied on examining the data and the results presented to assess the model's performance. Their feedback was as follows:

- **Accuracy of Anomaly Detection:** The experts confirmed that the anomalies identified by the autoencoder were indeed reflective of significant market events. For example, the high transaction volumes detected in specific months were aligned with known policy changes or major development projects in those periods.
- **Enhanced Efficiency:** The experts noted that the use of the autoencoder significantly improved the efficiency of their work by automating the detection of irregular patterns, which traditionally required extensive manual analysis. This allowed them to focus on deeper analysis and strategic planning.
- **Practical Utility:** The experts highlighted the practical utility of the model in providing early warnings of market shifts, enabling proactive decision making. The ability to detect anomalies in real time was particularly valued for its potential to mitigate risks and capitalise on emerging opportunities.

Overall, the results demonstrate the effectiveness of the autoencoder in identifying anomalies within the Kuwait Construction Market data. By leveraging the model's ability to learn from complex, high-dimensional data, we have shown that it can accurately pinpoint significant deviations indicative of underlying market dynamics. The validation by human experts, who assessed the same database we used and the results provided by us, further

underscores the model's practical applicability and its potential to enhance decision-making processes in the construction sector.

5.4. Discussion

This paper presents three main findings derived from the application of our trained autoencoder to analyse real estate volatility. First, it reveals that the heteroscedasticity commonly associated with real estate is not the principal challenge in generating highly volatile indices. Instead, fluctuations are more influenced by factors related to land price in indices exclusively focused on land transactions. Despite lacking many housing characteristics, land sales inherently exhibit high volatility, leading to more volatile indices for land-only transactions compared with those including house transactions, which show moderate volatility. This variation may also stem from property traders who frequently flip land or from the high transaction rates in certain cities.

Second, employing the autoencoder to strategically reduce the sample size and select specific sub-samples has significantly enhanced index performance in various aspects. Excluding data from regions mostly known for luxury and vacation homes markedly improves outcomes. Similarly, it seemed that the network focused on high-frequency cities, which, as a result, enhanced performance by mitigating the impacts of cities with low transaction frequencies, high volatility, and elevated costs.

Third, the autoencoder's capability to stratify data by city and long-term mean prices considerably bolsters the efficacy of central tendency measures used in index construction. While city stratification slightly edges out in performance, indices stratified by long-term mean prices hold competitive advantages. This approach aggregates a substantial volume of transactions per period, thereby minimizing the potential skewing effects of outliers, offering superior performance in certain scenarios compared with other stratification techniques.

Overall, based on our analysis, we found two main abnormalities in the data, which are the following:

1. **Transaction Volume of Lands:** The model identified significant outliers with volumes of 4278 and 13,211 transactions, which are substantially higher than typical volumes. These anomalies were correlated with specific market activities, such as new developments or policy changes.
2. **Transaction Volume of Houses:** Outliers included volumes of 1001 and 1150 transactions, also considerably higher than the average. These spikes were associated with specific events or shifts in market sentiment.

These outliers indicate special market activities in those specific cities. Abnormal behaviours, such as the extremely high transaction volumes in certain cities, seem to be driven by factors like new developments, policy changes, or other market influences, as stated earlier. The anomalies detected by the model highlight periods of unusual market activity, which could be critical for stakeholders in making informed decisions. For instance, the detection of abnormal spikes in transaction volumes can prompt further investigation into underlying causes such as economic changes, government policies, or new developments.

Based on our investigation on existing data, we can outline the following factors that might contribute to detected abnormalities:

- **Economic Changes:** Shifts in the economy, such as changes in interest rates, employment rates, impact of oil price change, and overall economic growth, have significantly impacted real estate transactions.
- **Government Policies:** The introduction of new government policies and incentives, such as subsidies for buyers and changes in zoning laws, has led to a dramatic change in transaction volumes.
- **Market Sentiment:** One of our hypotheses is that the general sentiment about the future of the property market might have caused fluctuations.

- **New Developments:** New developments and announcements of future developments in several districts has led to increased transactions as investors and homebuyers try to get in early.

5.5. Limitations

The limited availability of comprehensive property data and the challenges in creating reliable home price indexes in Kuwait highlight potential gaps in fully understanding market dynamics. While the current dataset provides a substantial foundation for analysis, generating and maintaining an optimal dataset is paramount for improving model accuracy and reliability. This lack of detailed data could complicate policy development, investment decisions, and economic forecasting related to the construction and real estate sectors. As usual, more extensive and detailed data would enhance the analysis, providing a clearer picture of market behaviours and facilitating more informed decision-making processes.

6. Conclusions

Through the analysis of anomalies identified by our autoencoder, we have acquired crucial insights into the most volatile aspects of the Kuwait Construction Market. This investigation not only has pinpointed areas requiring further research or strategic realignment but also has enabled us to project future market trends by examining past irregularities. The success of autoencoders in detecting these anomalies fundamentally relies on the quality and preprocessing of the input data, as well as precise tuning of the model's architecture and training parameters.

The anomalies highlighted by our analysis suggest critical areas for potential legislative action, market adaptation, or strategic modifications to address the prevailing challenges effectively. Understanding these anomalies requires a robust scientific approach that extends beyond mere detection. It involves a comprehensive exploration of their root causes, implications, and possible interventions, which should be pursued through further quantitative analyses and qualitative research.

Moreover, the stability of the construction market is influenced by internal factors, such as local material availability and government regulations, as well as external forces, including global oil prices and supply chain disruptions. Continuous monitoring of these elements will enhance our understanding of forthcoming market shifts and pinpoint vulnerabilities. This holistic approach ensures that strategic decisions are informed and responsive to both domestic and international dynamics, safeguarding the market's resilience and promoting sustainable growth.

While this paper primarily focuses on the methodology of using autoencoder neural networks for anomaly detection, the application to the specific dataset was limited by current data accessibility constraints. The Kuwait Construction Market dataset was selected due to its rich and diverse attributes, which are ideal for testing and validating advanced analytical models like autoencoders.

In future work, to enhance the robustness and applicability of our autoencoder model for anomaly detection in the Kuwait Construction Market, future research should prioritize the generation and maintenance of an optimal dataset. This involves comprehensive data coverage, high-quality data collection, and regular updates to capture the dynamic nature of the market. By addressing these aspects, we can significantly improve the model's performance and provide more accurate and actionable insights for stakeholders.

Author Contributions: Conceptualization, B.A.-S. and G.A.; methodology, B.A.-S.; software, B.A.-S.; validation, B.A.-S. and G.A.; formal analysis, B.A.-S.; writing—original draft preparation, B.A.-S. and G.A.; writing—review and editing, B.A.-S. and G.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the Public Authority for Civil Information and the Ministry of Justice.

Conflicts of Interest: Author Gholamreza Anbarjafari was employed by the company PwC FI—Advisory. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. AlSanad, S. Awareness, drivers, actions, and barriers of sustainable construction in Kuwait. *Procedia Eng.* **2015**, *118*, 969–983. [CrossRef]
2. Alfalah, A.A.; D’Arcy, E.; Stevenson, S. Constructing House Price Indices in an Emerging Market: The Case of Kuwait. *J. Real Estate Lit.* **2023**, *31*, 144–160. [CrossRef]
3. Alrasheed, K.; Soliman, E.; Al-Bader, H. Systematic Review of Construction Project Delays in Kuwait. *J. Eng. Res.* **2023**. [CrossRef]
4. Rezaeianjouybari, B.; Shang, Y. Deep learning for prognostics and health management: State of the art, challenges, and opportunities. *Measurement* **2020**, *163*, 107929. [CrossRef]
5. Baldi, P. Autoencoders, Unsupervised Learning, and Deep Architectures. In Proceedings of the ICML Workshop on Unsupervised and Transfer Learning, JMLR Workshop and Conference Proceedings. 2012. pp. 37–49. Available online: <https://proceedings.mlr.press/v27/baldi12a.html> (accessed on 9 July 2024)
6. Pinaya, W.H.L.; Vieira, S.; Garcia-Dias, R.; Mechelli, A. Autoencoders. In *Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 193–208.
7. Koushki, P.A.; Kartam, N. Impact of construction materials on project time and cost in Kuwait. *Eng. Constr. Archit. Manag.* **2004**, *11*, 126–132. [CrossRef]
8. Waly, A.F.; Thabet, W.Y. A virtual construction environment for preconstruction planning. *Autom. Constr.* **2003**, *12*, 139–154. [CrossRef]
9. Steinberg, F. Housing reconstruction and rehabilitation in Aceh and Nias, Indonesia—Rebuilding lives. *Habitat Int.* **2007**, *31*, 150–166. [CrossRef]
10. Al-Tabtabai, H.; Soliman, E. Oil prices drop effect on construction industry in Kuwait. *J. Eng. Res. (2307–1877)* **2022**, *10*, 1–15. [CrossRef]
11. Liu, F.; Umair, M.; Gao, J. Assessing oil price volatility co-movement with stock market volatility through quantile regression approach. *Resour. Policy* **2023**, *81*, 103375. [CrossRef]
12. Guo, C.; Zhang, X.; Iqbal, S. Does oil price volatility and financial expenditures of the oil industry influence energy generation intensity? Implications for clean energy acquisition. *J. Clean. Prod.* **2024**, *434*, 139907. [CrossRef]
13. Nawaz, A.; Khan, S.S.; Ahmad, A. Ensemble of Autoencoders for Anomaly Detection in Biomedical Data: A Narrative Review. *IEEE Access* **2024**, *12*, 17273–17289. [CrossRef]
14. Jarkas, A.M.; Horner, R.M.W. Creating a baseline for labour productivity of reinforced concrete building construction in Kuwait. *Constr. Manag. Econ.* **2015**, *33*, 625–639. [CrossRef]
15. Tong, G.K.; Ng, K.H.; Yap, W.S.; Khor, K.C. Construction of Optimal Stock Market Portfolios Using Outlier Detection Algorithm. In Proceedings of the Soft Computing in Data Science: 6th International Conference, SCDS 2021, Virtual Event, 2–3 November 2021; Proceedings 6; Springer: Berlin/Heidelberg, Germany, 2021; pp. 160–173.
16. Akinosho, T.D.; Oyedele, L.O.; Bilal, M.; Ajayi, A.O.; Delgado, M.D.; Akinade, O.O.; Ahmed, A.A. Deep learning in the construction industry: A review of present status and future innovations. *J. Build. Eng.* **2020**, *32*, 101827. [CrossRef]
17. Aslam, N.; Khan, I.U.; Alansari, A.; Alrammah, M.; Alghwairy, A.; Alqahtani, R.; Alqahtani, R.; Almushikes, M.; Hashim, M.A. Anomaly detection using explainable random forest for the prediction of undesirable events in oil wells. *Appl. Comput. Intell. Soft Comput.* **2022**, *2022*, 1558381. [CrossRef]
18. Ewubare, D.B.; Maeba, S.L. Effect of public expenditure in construction and transportation sectors on employment in Nigeria. *Int. J. Sci. Manag. Stud.* **2018**, *1*, 130–136.
19. AlSabah, R.; Refaat, O. Assessment of construction risks in public projects located in the state of Kuwait. *J. Eng. Res.* **2019**, *7*.
20. Alfalah, A.A. Challenges and Opportunities Facing Emerging Real Estate Markets: An Empirical Examination of the Kuwait Residential Real Estate Market. Ph.D. Thesis, University of Reading, Reading, UK, 2018.
21. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. Gpt-4 technical report. *arXiv* **2023**, arXiv:2303.08774.
22. Zhai, J.; Zhang, S.; Chen, J.; He, Q. Autoencoder and its various variants. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 415–419.
23. Chen, S.; Guo, W. Auto-encoders in deep learning—A review with new perspectives. *Mathematics* **2023**, *11*, 1777. [CrossRef]

24. Li, P.; Pei, Y.; Li, J. A comprehensive survey on design and application of autoencoder in deep learning. *Appl. Soft Comput.* **2023**, *138*, 110176. [[CrossRef](#)]
25. Singh, H. Anomaly Detection with Autoencoder. 2023. Available online: <https://github.com/AarnoStormborn/anomaly-detection-with-autoencoder> (accessed on 9 July 2024)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.