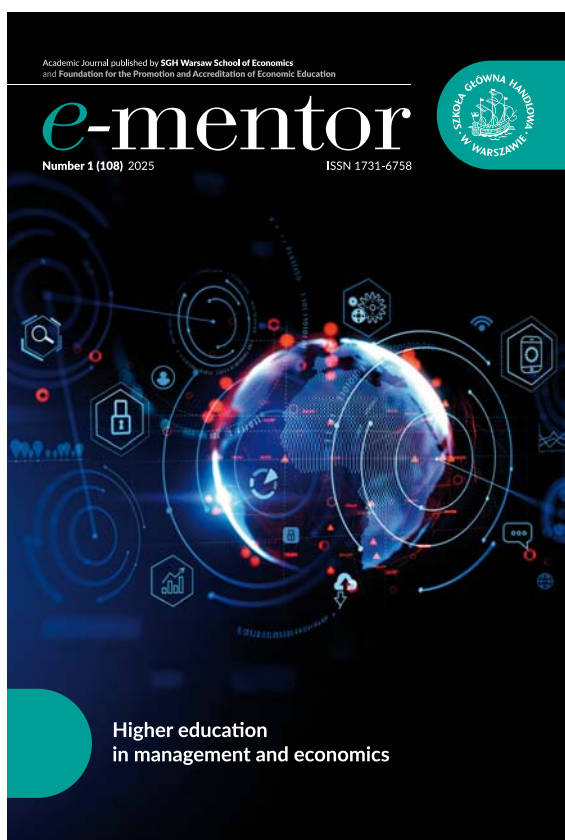


e-mentor

DWUMIESIĘCZNIK SZKOŁY GŁÓWNEJ HANDLOWEJ W WARSZAWIE
WSPÓŁWYDAWCA: FUNDACJA PROMOCJI I AKREDYTACJ KIERUNKÓW EKONOMICZNYCH

2025, nr 1 (108)



Pawlak, R., Wyrozębski, P., Pietras, I., & Parys, J. (2025). The Use of Machine Learning in Enhancing Data and Information Management Processes in the Context of Knowledge Management. *e-mentor*, 1(108), 52–62. <https://www.doi.org/10.15219/em108.1694>

Robert
Pawlak

Paweł
Wyrozębski

Ilona
Pietras

Joanna
Parys

The Use of Machine Learning in Enhancing Data and Information Management Processes in the Context of Knowledge Management

Abstract

The aim of this article is to describe how machine learning, including neural networks, is used in improving data and information management processes in the context of knowledge management in organisations. In the Polish literature, research on this topic tends to be fragmented. The empirical study presented in this article addresses the identified research gap concerning the use of machine learning methods in enhancing data and information management processes. The research makes use of critical literature analysis and the author's own research concerning the application of recurrent neural networks. The novel contribution made through this research in Poland was demonstration of the relationship between use of machine learning and the enhancement of knowledge management processes within organisations.


Keywords: data science, machine learning, neural networks, knowledge management, project management


Introduction


In an environment where information and technology solutions are increasingly important and data volumes are growing rapidly, enterprises need to continually adapt their data, information, and resulting knowledge management strategies to remain competitive in the markets. The application of advanced data mining and data science techniques, such as *Recurrent Neural Networks* (RNNs), is becoming not only a necessity but also a pivotal element of an organisation's success. The objective of this paper is to analyse how RNN models are used to forecast company stock prices, particularly the role of such models in enhancing data, information, and resulting knowledge management processes in organisations.

A critical literature review constitutes an essential element of this paper, providing an introduction to the research topic, making it possible to outline the theoretical context and formulate the main research objectives and questions. In the next step, the practical aspects of employing RNN models for stock price forecasting are discussed, illustrated using the example of Alphabet Inc. (Google). This analysis involved making use of the TensorFlow and Keras libraries, as well as the Python programming language, for implementing machine learning models, thereby enabling the forecasting of stock price trends based on historical data.

The digital era is driving the continuous growth of data volumes in organisations (Duque et al., 2022; Duque et al., 2023). With globalisation, competitiveness intensifies, compelling organisations to continuously adapt their management of the organisational knowledge that supports decision-making processes (Miller, 2014; Orad, 2020). Furthermore, data are an indispensable source for generating information and knowledge within enterprises (Chen et al., 2012; Report, 2018) – particularly in an environment where all aspects of the modern world, related to information and communication technologies,

Robert Pawlak, SGH Warsaw School of Economics, Poland,  <https://orcid.org/0000-0001-7228-4530>

Paweł Wyrozębski, SGH Warsaw School of Economics, Poland,  <https://orcid.org/0000-0002-6193-0055>

Ilona Pietras, SGH Warsaw School of Economics, Poland,  <https://orcid.org/0009-0000-0659-5162>

Joanna Parys, SWPS University, Poland,  <https://orcid.org/0009-0006-6453-5452>

contribute to their exponential growth. Data mining (Chen et al., 2012; Report, 2018) is a method for delving into and extracting data, as well as searching for knowledge in Business Intelligence systems, big data environments with large and variable data, and data warehouses. Patterns discovered during extraction and processing are regarded as information and knowledge, and should thus be viewed as representations of the same (Hevner & Chatterjee, 2010).

In a knowledge-based economy, there is a low capacity of enterprises to learn from undertaken projects, despite the wide availability of practical methods and tools (Ajmal et al., 2010; Paver & Duffield, 2019; Pawlak, 2021; Schindler & Eppler, 2003; Shergold, 2015; Williams, 2007; Wyrozębski, 2014). Many organisations lack the ability to learn from historical data, which forces them to reinvent the wheel (Paver & Duffield, 2019). Knowledge management processes in companies involve collecting information, data, and experiences (APM, 2019; OGC, 2017; PMI, 2017), but questions about the results delivered and the effectiveness of the process are continually being examined in research in the fields of management and quality sciences (Paver & Duffield, 2019; Wiewiora & Murphy, 2015; Williams, 2008; Wyrozębski, 2012). Paver and Duffield (2019) claim that the fundamental challenge lies in the appropriate and correct execution of actions in response to accumulated knowledge. They suggest that the answers can be found in the data collected by organisations and our ability to interpret them, rather than in the application of a specific procedure or method.

Knowledge management, much like the concept of knowledge itself, can be defined in many different ways. The absence of a cohesive description of knowledge management prompts further scholarly reflection. According to Wiig (1997), knowledge management is the systematic, explicit, and deliberate construction, renewal, and utilisation of knowledge to maximise organisational performance. In contrast, Probst et al. (2002) view it as a collection of practices that an organisation employs to create, store, utilise, and share knowledge.

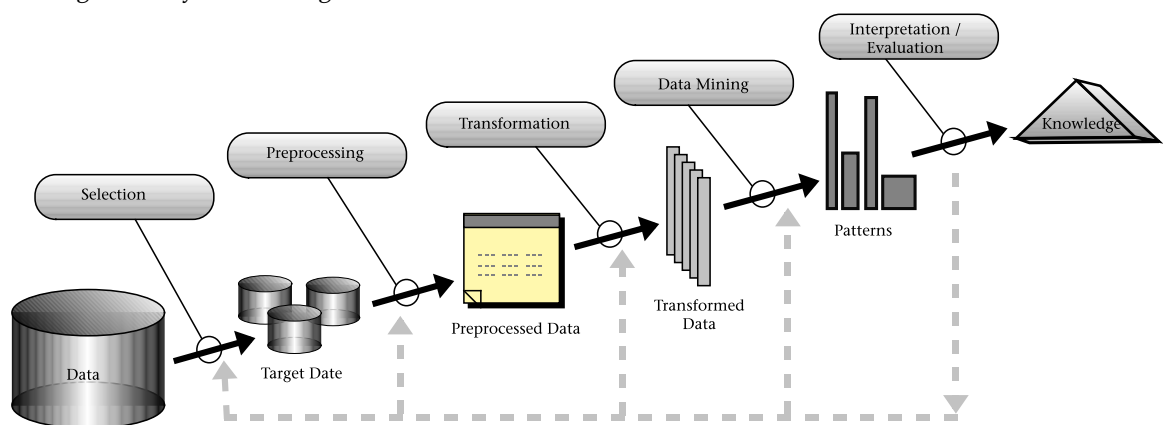
Kotnour and Vergopia (2005) are of the opinion that it is a set of proactive activities that aid the organisation in the creation, acquisition, dissemination, and utilisation of knowledge. The definition that best aligns with the authors' perspective is that of Romanowska, who describes knowledge management knowledge as a method of managing an enterprise that encompasses the entirety of processes enabling the creation, dissemination, and utilisation of knowledge to achieve the enterprise's objectives (Romanowska, 2004, p. 702).

In the case study described in this article, it was possible to examine the subsequent phases of the knowledge discovery process. The gathered data provided information for a machine learning process implemented using neural networks, illustrating how enterprises can be supported in data and information management processes in the context of organisational knowledge management (Drucker, 1995; Łobejko, 2004; Nazarko, 2018).

Theoretical Aspects of Using Data Mining and Data Science Solutions in Data and Information Management Processes in Organisations

In order to identify elements of the process related to data mining that are critical for data and information management in organisations, a review of databases such as Scopus, Science Direct, Web of Science, and Elsevier was conducted. As a result, a single, coherent definition of data mining was established. The meaning of this phrase, most commonly translated into Polish as *eksploracja danych*, is described as a process leading to the analysis and use of large datasets to extract valuable insights and the resulting knowledge (Duque et al., 2023; Fayyad et al., 1996). There are five main phases of knowledge discovery in organisations (Fayyad et al., 1996): selection, preprocessing, transformation, data mining, and evaluation and interpretation (figure 1).

Figure 1
Knowledge Discovery Process in Organisational Databases



Source: "From data mining to knowledge discovery in databases", U. Fayyad, G. Piatetsky-Shapiro, & P. Smyth, 1996, *AI Magazine*, 17(3), p. 37 (<https://doi.org/10.1609/aimag.v17i3.1230>).

According to Fayyad et al. (1996) and Maimon and Rokach (2010), the process of knowledge discovery consists of five main stages. Selection involves defining the objectives and choosing the data to be analysed in the knowledge discovery process. Preprocessing, in turn, enhances the reliability of data through processes such as filtering and cleaning. The transformation element enables the development of the most effective data model to ensure high-quality feedback. Data mining facilitates the discovery of trends and recurring patterns within the data. The final stage, involving evaluation and interpretation, serves to uncover new knowledge and/or update the process if necessary.

The data mining process is a vital component of the holistic data orchestration framework within organisations, as depicted in figure 2. Data orchestration is portrayed in the literature (Schwe, 2023) as a process that outlines the sequence of all activities related to data management within organisations. Data, information, and the resultant knowledge become pivotal resources for enterprises, upon which they base their decision-making in business operations. In such an environment, data orchestration processes are the prevailing method of managing data flow within an organisation, enabling decision-making based for example on the predictive outputs of machine learning models (Schwe, 2023).

The initial stage of the data orchestration process is dedicated to data ingestion, and involves the collection, importing, and processing of data from various sources for further analysis, storage, or presentation. Data can be extracted using methods such as Extract,

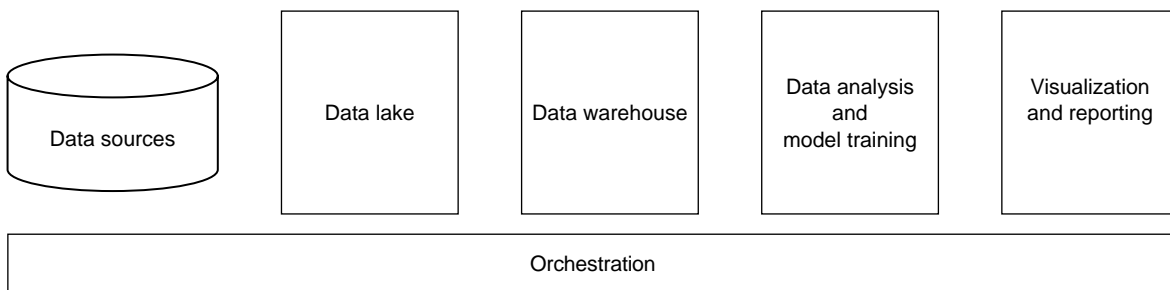
Transform, Load (ETL), which involve obtaining and then transformation and loading of the data into the target system. Data can be ingested in real time (streaming) or through batch processing.

As illustrated in figure 2, data can be stored in a data lake or data warehouse depending on their format, namely structured, unstructured, or semi-structured. Data lakes are a type of architecture that facilitate the collection, storage, and analysis of diverse data types in their original format. Data warehouses, on the other hand, are digital storage systems that integrate and harmonise large volumes of data from a variety of sources.

During the data analysis and modelling stage of the process, the solution most frequently referenced in scholarly literature on the subject of data science is the Machine Learning Operations (MLOps) model, which is a collection of principles and practices that define data management methods within machine learning processes and can serve as one of the key elements in enhancing data and knowledge management processes within organisations (Schwe, 2023).

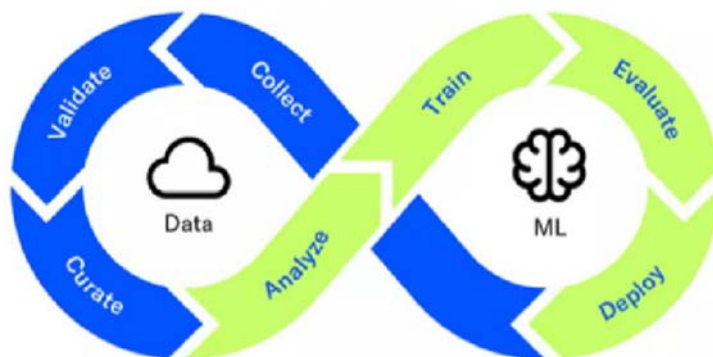
As demonstrated in figure 3, this model comprises seven primary steps. These are collecting, validating, and curating data, followed by analysing the data, training the model, evaluating the obtained outcomes, and finally deploying the model within the organisation. These MLOps model steps evidently form a loop and result in a recursive process, which organisations internalise within their business operations, thereby refining their data, information, and resulting knowledge management processes (Schwe, 2023).

Figure 2
Data Orchestration Process within Organisations



Source: authors' own work.

Figure 3
Machine Learning Operations (MLOps) Model



Source: authors' own work.

The Use of Machine Learning in Enhancing Data...

In the final stage of the data orchestration process, the practice of discovering patterns and making decisions based on them is accompanied by the emergence of trend lines, cycles, and/or seasonality (Nazarko, 2018). The aim of data mining and orchestration processes is to identify unexpected dependencies and patterns (Adeniyi et al., 2016; Duque et al., 2023). The objectives of data processing as part of data mining can be divided into two main categories and further into subcategories, as illustrated in figure 4.

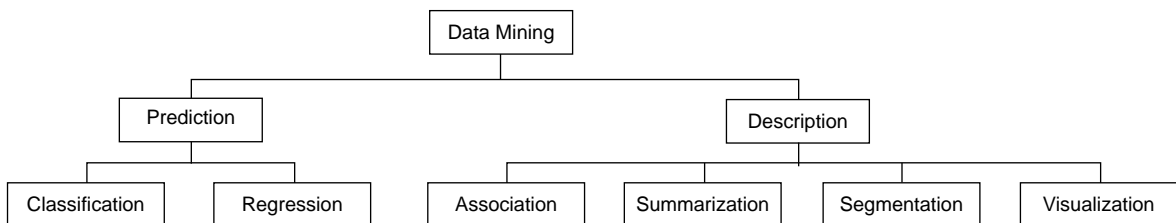
In figure 4, two subcategories for prediction analysis are presented, namely classification and regression. Conversely, the description category comprises the following subcategories: association, summarisation, segmentation, and visualisation (Adeniyi et al., 2016; Duque et al., 2023). Classification is the process of assigning classes to objects based on their attributes. Regression enables the attribution of specific values to various variables, e.g. predicting the probability of surviving a disease based on the results of diagnostic tests. Association, in turn, is a means of exploring relationships between variables, for example, through the placement of products together in a shopping cart (commonly known as *market basket analysis*). Summarisation is the most advanced method, stem-

ming from additive rules, visualisation techniques, and the discovery of relationships between variables. Segmentation (also referred to as *clustering*) allows for heterogeneous division of a population into several homogeneous segments, where data are grouped according to similarity. The results of data mining are presented using charts and diagrams (a subcategory of visualisation), which enable the identification and presentation of patterns and trends in the analysed data in graphic format (Easterby-Smith & Lyles, 2003).

One of the approaches to knowledge management in organisations is presented by Easterby-Smith and Lyles (2003), who describe the interrelations between organisational learning, organisational knowledge, learning organisation and knowledge management from the perspectives of process and content, as well as theory and practice. Figure 5 illustrates the interconnections and mapping of these four related terms.

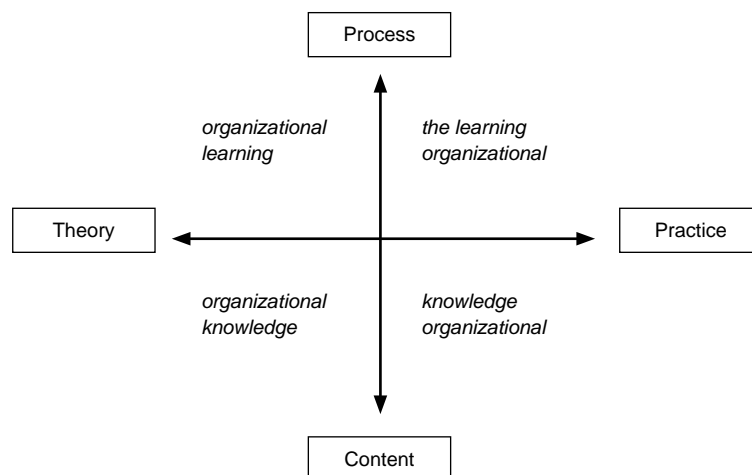
Initiatives related to developing, disseminating, and utilising knowledge help organisations internalise the acquired knowledge, leading to improvements in operational efficiency. In turn Dixon (1994), describing the organisational learning cycle, suggested that ac-

Figure 4
Categories and Subcategories in Data Mining Processes



Source: "Data mining applied to knowledge management", J. Duque, F. Silva, & A. Godinho, 2023, *Procedia Computer Science*, 219, p. 456 (<https://doi.org/10.1016/j.procs.2023.01.312>).

Figure 5
Interdependencies Between Organisational Learning, Learning Organisation, Organisational Knowledge, and Knowledge Management



Source: *The Blackwell handbook of organizational learning and knowledge management* (p. 17), M. Easterby-Smith & M. Lyles, 2003, Blackwell.

cumulated knowledge holds less significance than the processes required for its ongoing pursuit and creation. The described processes are directly related to Nonaka's Socialisation, Externalisation, Combination, Internalisation (SECI) model (Dixon, 1994).

An analysis of these considerations might lead to the research question of how data mining and data science solutions enhance data and information management processes in the context of knowledge management in organisations? This topic is widely discussed in the literature (Duque, 2023; Jifa & Lingling, 2018; Raschka & Mirjalili, 2021; Shaeffer & Makatsaria, 2021), particularly regarding the increasing interest in the use of machine learning in data and information management processes in organisations.

Using Machine Learning in Knowledge Management in Organisations

When exploring the connections between definitions of concepts such as data, Data, Information, Knowledge, Wisdom (DIKW), big data, and data science, the model of knowledge creation in organisations discussed in the literature is also relevant (Jifa & Lingling, 2018) (figure 6).

DIKW is explained further in literature (Ackoff, 1989). *Data* is defined as the most fundamental model, followed by *information*, which provides context to the data. The next level is *knowledge*, acquired by organisations based on the collected information, while *wisdom* is meant as a guide to when and why the acquired knowledge should be applied. Within the DIKW model, data is a fundamental fact and a raw material, which will only become useful if we allow it to evolve into information, knowledge, and wisdom. At the data and information levels, various methods can be employed, such as data mining, text mining, web mining, and specialised tools including databases, data marts, and management information systems (Leondes, 2010). In the progression towards the third level, namely *knowledge*, organisations implement Knowledge Discovery in Databases (KDD) solutions and knowledge engineering methods

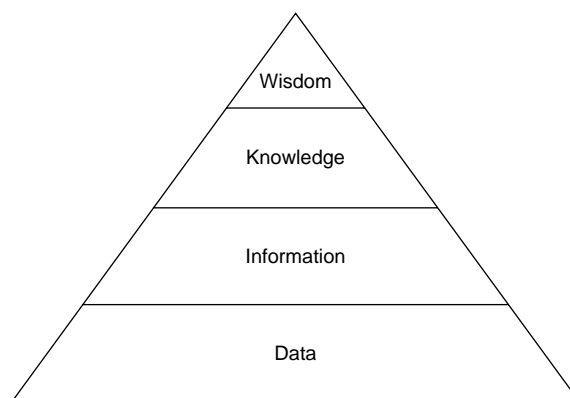
(Leondes, 2010). In intelligent knowledge management processes, organisations use what is referred to as *domain-driven data mining* (Cao and Zhang, 2006). This is a method that involves generating expert knowledge, termed an *intelligent knowledge base* (Zhang, 2009).

Organisations implementing machine learning models (machine learning is a subset of artificial intelligence), are undergoing a transformational revolution in knowledge management practices (Dalkir, 2017; Larose & Larose, 2015; Schwe, 2023). Machine learning algorithms excel at processing large volumes of data both quickly and accurately. By leveraging ML, organisations can uncover previously unknown patterns within the collected data. A neural network-based analytics approach facilitates a deeper understanding of complex relationships among the variables under investigation and can encourage more informed decision-making.

In the field of intelligent knowledge discovery, machine learning algorithms can also promote discovering new knowledge, for example by automating information retrieval processes. *Natural Language Processing* (NLP) techniques enable intelligent search systems to understand and interpret the context, intent, and sentiment behind the queries being analysed. Automated knowledge capture using machine learning algorithms also helps to enhance the discovery of trends and patterns in data obtained from distributed source systems (Dalkir, 2017; Larose & Larose, 2015; Schwe, 2023).

Conversely, intelligent content recommendation techniques (Adeniyi et al., 2016) – available through neural networks – enable the provision of personalised content recommendations based on the preferences of the enterprises' customers, their interests, and past interactions (e.g. purchasing preferences). In such a scenario, ML algorithms help to predict future customer behaviour in a rapidly changing market, enabling businesses to anticipate potential challenges and seize emerging opportunities (Dalkir, 2017; Larose & Larose, 2015; Schwe, 2023).

Figure 6
Data, Information, Knowledge, Wisdom (DIKW) Model



Source: "Data, DIKW, Big Data and Data Science", G. Jifa & Z. Lingling, 2018, *Procedia Computer Science*, 31, p. 815 (<http://dx.doi.org/10.1016/j.procs.2014.05.332>).

Case Study: Using a Recurrent Neural Network to Forecast Company Stock Prices

In the theoretical framework of methodological assumptions underlying neural network models, it is possible to predict stock prices based on historical data (Li et al., 2018; Raschka & Mirjalili, 2021; Saud & Shakya, 2020; Zhu, 2020; Zhu et al., 2022). Nevertheless, it is impossible to precisely determine all the factors that may influence stock prices and how these factors will impact the stock markets. One reason for this is the assumption that a forecasting model should be capable of addressing non-linear problems, and stock price forecasting is strongly non-linear. The use of RNNs is appropriate for forecasting because stock prices exhibit time series characteristics (Li et al., 2018; Raschka & Mirjalili, 2021; Saud & Shakya, 2020; Zhu, 2020; Zhu et al., 2022), and the purpose of RNNs is to process sequential data.

The literature (Zhu et al., 2022) provides examples of neural network models such as *Long Short-Term Memory* (LSTM) used to predict stock prices based on closing prices. The LSTM model proposed by Hochreiter and Schmidhuber (1997) was modified based on the RNN structure, which addresses the limitation of RNN models in representing the long-term memory of time series (Staudemeyer & Morris, 2019). The application of this modified LSTM model – which can pass data to each layer – ensures the existence of short-term memory during the training of long-term memory. Furthermore, LSTM also mitigates the problem of vanishing gradients that occurs in recurrent neural networks when processing long sequences of data (Hochreiter & Schmidhuber, 1997). In LSTM models, short-term

memory is responsible for retaining information from the most recent time steps, enabling the model to swiftly adapt to current changes in the data.

In contrast, long-term memory accumulates significant patterns from prior data, assisting the model in understanding and predicting long-term dependencies and trends. In LSTM, short-term memory is managed by the cell state, which retains current information between time steps, whereas long-term memory uses a gating mechanism. This enables the LSTM to model both short-term dependencies and long-term patterns while mitigating the vanishing gradient effect.

Using RNNs to forecast corporate stock prices is a common and recommended approach within the examined body of literature (Raschka & Mirjalili, 2021; Saud & Shakya, 2020; Zhu, 2020; Zhu et al., 2022). TensorFlow and Keras libraries enable the application of deep learning mechanisms, which help to speed up the processing of machine learning tasks.

The analyses conducted for the purposes of this research, using the Python programming language, were performed using Anaconda package version 2.3.2. Jupyter Notebook software suite version 6.5.2, provided within the Anaconda package, served as the *Integrated Development Environment* (IDE). In the initial step of data preparation, the Pandas, Numpy, and Matplotlib libraries were used, as demonstrated in figure 7. Publicly available data on Google's (Alphabet Inc) stock prices from the Kaggle platform was downloaded and used (Google Stock Price, n.d.).

As presented in figure 7, the data retrieved for analysis consists of Google's (Alphabet Inc.) stock prices from 2012 to 2016, specifically the opening and closing prices as well as the highest and lowest

Figure 7
Python Scripts – Part 1.
Data Processing

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

In [2]: dataset_train = pd.read_csv('Google_Stock_Price_Train.csv')
training_set = dataset_train.iloc[:, 1:2].values

In [3]: print(dataset_train)

      Date  Open  High  Low  Close  Volume
0  1/3/2012  325.25  332.83  324.97  663.59  7,380,500
1  1/4/2012  331.27  333.87  329.08  666.45  5,749,400
2  1/5/2012  329.83  330.75  326.89  657.21  6,590,300
3  1/6/2012  328.34  328.77  323.68  648.24  5,405,900
4  1/9/2012  322.04  322.29  309.46  620.76  11,688,800
...
1253 12/23/2016  790.90  792.74  787.28  789.91  623,400
1254 12/27/2016  790.68  797.86  787.66  791.55  789,100
1255 12/28/2016  793.70  794.23  783.20  785.05  1,153,800
1256 12/29/2016  783.33  785.93  778.92  782.79  744,300
1257 12/30/2016  782.75  782.78  770.41  771.82  1,770,000

[1258 rows x 6 columns]

In [4]: print(training_set)

[[325.25]
 [331.27]
 [329.83]
 ...
 [793.7 ]
 [783.33]
 [782.75]]
```

Source: authors' own work.

prices, alongside transaction volumes for each respective day. For the purposes of the present analyses, two data sets were collected: one stock price set served as the training set, while the second contained test data. The aim of the analyses was to forecast Google's stock price trend for the year 2017 and subsequently compare it with the dataset of real-world data (comprising verified data for that period) to evaluate the quality of the outcomes (forecasts). According to the predictions of the Brownian motion theory (Heller, 2023; Mandelbrot & Taylor, 1967), future stock price variations are independent of past data, indicating the non-linear nature of the relationship between the variables. Thus, in line with the predictions based on this theory, stock prices cannot be predicted accurately, and only trends (rise or fall) can be foreseen. For the purpose of this analysis, and to also minimise the complexity of the study, only the opening stock price was used for forecasting.

The next step in the analyses was to map the stock prices to the range (0, 1) to prevent overfitting of the model (particularly when using the sigmoid activation function¹), as illustrated in figure 8 (see Appendix 1). The `MinMaxScaler` method from the Scikit-learn (Sklearn) library was used for the purposes of this normalisation².

In the subsequent step, a dedicated data structure encompassing 60 time stamps was prepared, based on which the RNN would predict the 61st stock price of the enterprise under analysis. The number of preceding time stamps was determined to be 60 on the basis of previous experimental observations (Zhu et al., 2022). As a result of these activities, the variable `X_train` was transformed into a nested list containing lists of prices spanning 60 time stamps. Conversely, the variable `y_train` is a list of stock prices containing the prices for the following day that correspond to the variables within the `X_train` list.

Figure 9 illustrates a section of the `X_train` and `y_train` set of variables (see Appendix 1). Each row in the `X_train` set, containing 60 rescaled stock prices, is subsequently used to predict the corresponding next day's opening stock price within the `y_train` outcome set.

During programming, three LSTM layers were built as an initial step. A Dropout layer with a rate of 0.2 was inserted between the three prepared layers, to minimise model overfitting. The fourth LSTM layer is a fully connected layer, comprised of a single neuron, designed to predict the future price of a single share of Google. To estimate parameters, the *Adam* optimiser was used. The maximum number of iterations was set to 50, and the *Mean Squared Error* (MSE) method was employed to minimise loss until a value

of zero was reached (see figure 10 in Appendix 1). The authors of this article recommend considering, for similar analyses, the *Mean Absolute Error* (MAE) function, which is less sensitive to large errors, or the *Huber Loss* function, a combination of MSE and MAE, to better manage outliers. In the authors' view, both error functions can help to formulate more stable and robust forecasts in time series models.

In these analyses, the primary model used for predicting continuous values in the LSTM model is a neural network regressor. In the first step, we initialise the model using the *Sequential* function. Next, we add LSTM layers and configure the optimiser and loss function, which allow the model training to commence. The *Units* parameter specifies the number of LSTM neurons in the layer. A count of 50 neurons provides high model dimensionality, a means of forecasting upward and downward stock price trends. Setting the *Return Sequences* parameter to *True* enables the addition of another LSTM layer to the previously created layers. The *Input Shape* parameter corresponds to the number of time stamps and the number of indicators. In turn, the *Dropout* parameter – with a value set to 0.2 – implies that 20% of the 50 neurons will be randomly omitted during each training iteration (as illustrated in figure 11 – see Appendix 1).

In the subsequent step, the RNN was compiled using the *Stochastic Gradient Descent* (SGD) algorithm³, along with the loss function and optimiser previously described. Then, the model training process was initiated, aimed at forecasting the opening stock prices of the analysed company.

The network is fitted to the training set over 100 epochs – see figure 12 in Appendix 1. Its weights are updated for every 32 stock prices. The batch size is set to 32.

As a result of the training process, the loss value reached was 0.0018. In the subsequent step, test data is retrieved to conduct forecasts and, ultimately, to present a visualisation of the results (figures 13 and 14 in Appendix 1).

The training and test datasets are combined for the forecasting process since we use closing prices from the previous 60 days to predict the next day's price. In other words, we need stock prices from the 60 days preceding the first date in the test dataset, as illustrated in figure 14 (see Appendix 1).

The final step is the visualisation of the results of the analysis, to provide a graphic representation of the final findings to as an aid for interpretation (figure 15).

As illustrated in figure 15, the predicted stock price curve diverges from the real-world values because the model cannot adequately respond to non-linear changes in real-world variables. Nonetheless, the

¹ The sigmoid activation function is a concept used in artificial intelligence to define the function by which the output value of neural network neurons is calculated.

² Scikit-learn is a Python library that helps implement machine learning algorithms, such as regression, classification, clustering, model evaluation, and numerous others.

³ *Stochastic Gradient Descent* (SGD) is an optimisation algorithm used to minimise the cost function. The cost function serves to measure how closely the model's predictions align with the real-world values.

Figure 15

Python Script – Visualisation of Results

```
In [26]: plt.plot(real_stock_price, color = 'red', label = 'Real Google Stock Price')
plt.plot(predicted_stock_price, color = 'blue', label = 'Predicted Google Stock Price')
plt.title('Google Stock Price Prediction')
plt.xlabel('Time')
plt.ylabel('Price')
plt.legend()
plt.show()
```



Source: authors' own work.

model responds correctly to smooth transformations. The stock price forecasts presented in figure 15, which exhibit a discontinuous nature, have a detrimental effect on the model, resulting in deviations from real-world stock prices. The smooth nature of changes contributes to the accuracy of stock price forecasts generated by the model – as demonstrated in the analysed example – appropriately reflecting upward and downward trends in stock prices.

Summary and Discussion of Research Results

When attempting to interpret the obtained analysis results, one must agree with the referenced literature (Balicka, 2023; Chen et al., 2012; Drucker, 1995; Duhon & Elias, 2008; Duque et al., 2023; Jifa & Lingling, 2018; Łobejko, 2004; Milton, 2010; Raschka & Mirjalili, 2021; Report, 2018; Saud & Shakya, 2020; Schindler & Ep-

pler, 2003; Shaeffer & Makatsaria, 2021; Wiewiora & Murphy, 2015; Williams, 2007; Williams, 2008; Wyrozębski, 2014; Zedwitz, 2002; Zhu, 2020; Zhu et al., 2022) that exploring the application of machine learning methods to enhance data and information management processes in the context of knowledge management within organisations is of particular interest from a research perspective, especially in the era of the digital revolution and the exponentially increasing volume of data available in organisations.

In the example cited from the literature (Zhu et al., 2022), the closing prices of company stocks were analysed, whereas for the purposes of analysis in the present study, the opening price of Google stocks was chosen as the forecast price. Similar results were obtained in both instances, indicating that RNNs employing the LSTM model can accurately predict stock price trends. However, they are unable to forecast the precise stock price for a specific date (figure 16).

Figure 16

Comparison of Forecast Stock Prices with Real-World Values (20-year period)



Source: "Predicting Google's Stock Price with LSTM Model", T. Zhu, Y. Liao & Z. Tao, 2022, *Proceedings of Business and Economic Studies*, 5(5), p. 85 (<http://dx.doi.org/10.26689/pbes.v5i5.4361>).

Addressing the research question posed in this paper regarding the effectiveness/efficacy of the selected model, the observation cited in the literature proved to be correct, that only the trend line (either increase or decrease in stock prices) can be accurately forecast. This is primarily due to the fact that the RNN model employing the LSTM method cannot react swiftly to non-linear changes in the real-world variable. In accordance with the theory of Brownian motion (Mandelbrot & Taylor, 1967) – as discussed in this paper – it is not possible to accurately reflect future stock price changes solely based on historical data. However, the model does accurately forecast upward and downward trends in the stock prices of the analysed company.

Zhu (2020) compared RNN models with classical regression models and simple predictive methods in the context of the task of forecasting stock prices. He demonstrated that RNNs outperform linear models in detecting non-linear patterns in stock market data, although RNNs encounter difficulties in accurately predicting sudden changes in stock prices. In contrast, Saud and Shakya (2020) compared the effectiveness of RNN, LSTM, and simple autoregressive models. The results of their analyses revealed that deep learning models, such as LSTM, handle non-linear dependencies more effectively, while simpler autoregressive models are more efficient in terms of time and resources, which can be advantageous for shorter prediction horizons. In their study, Zhu et al. (2022) compared the LSTM with traditional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA), demonstrating clear advantages of the LSTM in detecting long-term trends and patterns, although in the short term, ARIMA offered similar effectiveness at a lower computational cost.

Using various machine learning techniques described in the present study – ranging from knowledge discovery and capture algorithms to intelligent content recommendation, and predictive analytics facilitating decision-making – can enhance data and information management processes in the context of knowledge management and lead to more efficient dissemination and assimilation within companies endeavouring to employ machine learning methods. Improved data and information management processes in the context of knowledge management are part of the suite of data orchestration processes within organisations. Such solutions include processes aimed at acquiring data from numerous distributed data sources. Enhancements in such data acquisition processes can help to improve the data quality and – subsequently – the accurate identification of appropriate source systems. Another process that can be optimised is the data analysis and modelling process. Using machine learning techniques can enhance the quality of analyses and the resulting output data.

The conclusions drawn from the analyses conducted in this article – in light of the literature review – confirm the increasing significance of data mining and data science solutions in data and information

management processes in the context of knowledge management in organisations. Moreover, it has been demonstrated that the techniques and methods of data mining and data science constitute an integral part of organisational activities and facilitate organisational learning in the processes that assimilate new knowledge into business operations. Through the presentation of practical applications of the chosen machine learning model – as an example of use of data mining and data science solutions – it was possible to illustrate how the presented model helps to improve knowledge management processes in organisations, while simultaneously enabling an in-depth exploration of data and information.

When indicating potential directions for future research, the use of GRU, Vanilla, and Convolutional Neural Networks should be recommended as alternatives to the LSTM (GitHub, 2024), for forecasting stock prices based on historical data. Including more than just the opening or closing prices of a stock in the analysis is also advisable. The attempt to employ different machine learning techniques could help to enhance and broaden knowledge in the context of improving data and information management processes in learning organisations.

The appendix is available in the online version of the journal.

References

- Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
- Adeniyi, D., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, 12(1), 90–108. <https://doi.org/10.1016/j.aci.2014.10.001>
- Ajmal, M., Helo, P., & Kekäle, T. (2010). Critical factors for knowledge management in project business. *Journal of Knowledge Management*, 14(1), 156–68. <http://dx.doi.org/10.1108/13673271011015633>
- APM. (2019). *APM body of knowledge* (7th ed.). Association for Project Management.
- Balicka, H. (2023). Internet rzeczy i modele uczenia głębokiego w zrównoważonym rozwoju inteligentnych miast. *Współczesna Gospodarka*, 16(1), 27–43. <https://doi.org/10.26881/wg.2023.1.03>
- Cao, L., & Zhang, C. (2006). Domain-driven actionable knowledge discovery in the real world. In W. K. Ng, M. Kitsuregawa, J. Li, & K. Chang (Eds.), *Advances in knowledge discovery and data mining* (pp. 821–830). PAKDD. Lecture Notes in Computer Science 3918. https://doi.org/10.1007/11731139_96
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188. <https://doi.org/10.2307/41703503>
- Dalkir, K. (2017). *Knowledge management in theory and practice* (3th ed.). MIT Press.
- Dixon, N. (1994). *The organizational learning cycle: How we can learn collectively*. McGraw-Hill.

The Use of Machine Learning in Enhancing Data...

- Drucker, P. (1995). *Zarządzanie w czasach burzliwych*. Czytelnik.
- Duhon, H., & Elias, J. (2008). Why it is difficult to learn lessons: insights from decision theory and cognitive science. *SPE Project, Facilities & Construction*, 3(3), 1–7. <http://dx.doi.org/10.2118/110211-MS>
- Duque, J., Godinho, A., & Vasconcelos, J. (2022). Knowledge data extraction for business intelligence. A design science research approach. *Procedia Computer Science*, 204, 131–139. <https://doi.org/10.1016/j.procs.2022.08.016>
- Duque, J., Silva, F., & Godinho, A. (2023). Data mining applied to knowledge management. *Procedia Computer Science*, 219, 455–461. <https://doi.org/10.1016/j.procs.2023.01.312>
- Easterby-Smith, M., & Lyles, M. (Eds.). (2003). *The Blackwell handbook of organizational learning and knowledge management*. Blackwell.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54. <https://doi.org/10.1609/aimag.v17i3.1230>
- GitHub. (n.d.). Retrieved July 7, 2023, from <https://github.com/huseinzol05/Stock-Prediction-Models>
- Google Stock Price. (n.d.). Retrieved December 20, 2023, from <https://www.kaggle.com/datasets/vaibhavsn/google-stock-prices-training-and-test-data/data>
- Heller, M. (2023). *Filozofia przypadku* (7th ed.). Copernicus Center Press.
- Hevner, A., & Chatterjee, S. (2010). *Design research in information systems theory and practice*. Springer.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>
- Jifa, G., & Lingling, Z. (2018). Data, DIKW, Big Data and Data Science. *Procedia Computer Science*, 31, 812–821. <http://dx.doi.org/10.1016/j.procs.2014.05.332>
- Kotnour, T., & Vergopia, C. (2005). Learning-based project reviews: observations and lessons learned from the Kennedy Space Center. *Engineering Management Journal*, 17(4), 30–38. <http://dx.doi.org/10.1080/10429247.2005.11431670>
- Larose, D., & Larose, C. (2015). *Data mining and predictive analytics* (2nd ed.). John Wiley & Sons.
- Leondes, C. T. (2010). *Intelligent knowledge-based systems: Business and technology in the new millennium*. Springer.
- Li, S., Li, W., Cook, C., Zhu, C., & Gao, Y. (2018). Independently Recurrent Neural Network (IndRNN): Building a Longer and Deeper RNN. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5457–5466). IEEE. <https://doi.org/10.1109/CVPR.2018.00572>
- Łobejko, S. (2004). *Systemy informacyjne w zarządzaniu wiedzą i innowacją w przedsiębiorstwie*. Szkoła Główna Handlowa w Warszawie.
- Maimon, O., & Rokach, L. (2010). Introduction to knowledge discovery and data mining. In M. Oded, & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (2nd ed., pp. 1–15). Springer. https://doi.org/10.1007/978-0-387-09823-4_1
- Mandelbrot, B., & Taylor, H. (1967). On the distribution of stock price differences. *Operations Research*, 15(6), 1057–1062. <http://dx.doi.org/10.1287/opre.15.6.1057>
- Miller, R. (2014, May 22). *Actually, every company is a big data company*. <https://techcrunch.com/2014/05/22/actually-every-company-is-a-big-data-company/>
- Milton, N. (2010). *The lessons learned handbook: Practical approaches to learning from experience*. Elsevier.
- Nazarko, J. (Ed.). (2018). *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. IV. Prognozowanie na podstawie modeli trendu*. Oficyna Wydawnictw Politechniki Białostockiej.
- OGC. (2017). *Managing successful projects with PRINCE2* (6th ed.). Office of Government Commerce.
- Orad, A. (2020, February 14). Why every company is a data company. *Forbes*. <https://www.forbes.com/sites/forbestechcouncil/2020/02/14/why-every-company-is-a-data-company/>
- Paver, M., & Duffield, S. (2019). Project management lessons learned: The elephant in the room. *Journal of Modern Project Management*, 6(3), 105–121.
- Pawlak, R. (2021). *Znaczenie i rola doświadczeń projektowych w zarządzaniu wiedzą w projektach* [Doctoral dissertation]. Szkoła Główna Handlowa w Warszawie. <https://sgh.bip.gov.pl/articles/view/775774>
- PMI. (2017). *A guide to the project management body of knowledge (PMBOK Guide)* (6th ed.). Project Management Institute.
- Probst, G., Raub, S., & Romhardt, K. (2002). *Zarządzanie wiedzą w organizacji*. Oficyna Ekonomiczna.
- Raschka, S., & Mirjalili, V. (2021). *Python machine learning and deep learning* (3th ed.). Helion.
- Report. (2018). *The 2018 global data management benchmark report*. Experian Information Solutions, Inc. <https://www.readkong.com/page/the-2018-global-data-management-benchmark-report-4007555>
- Romanowska, M. (Ed.). (2004). Zarządzanie wiedzą. In *Leksykon zarządzania* (p. 702). Difin.
- Saud, A., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *Procedia Computer Science*, 167, 788–798. <https://doi.org/10.1016/j.procs.2020.03.419>
- Shaeffer, C., & Makatsaria, A. (2021). Framework of data analytics and integrating knowledge management. *International Journal of Intelligent Networks*, 2, 156–165. <https://doi.org/10.1016/j.ijin.2021.09.004>
- Schindler, M., & Eppler, M. J. (2003). Harvesting project knowledge: a review of project learning methods and success factors. *International Journal of Project Management*, 21(3), 219–228. [https://doi.org/10.1016/S0263-7863\(02\)00096-0](https://doi.org/10.1016/S0263-7863(02)00096-0)
- Schwe, M. (2023, August 22). *Data orchestration: A conductor for the modern data platform*. <https://medium.com/@mikesheve/19587/data-orchestration-682af6919898>
- Shergold, P. (2015). *Learning from Failure: Why large government policy initiatives have gone so badly wrong in the past and how the chances of success in the future can be improved*. Western Sydney University.
- Staudemeyer, R. C., & Morris, E. R. (2019). *Understanding LSTM: A tutorial into long short-term memory recurrent neural networks*. arXiv:1909.09586. <https://doi.org/10.48550/arXiv.1909.09586>
- Wiewiora, A., & Murphy, G. (2015). Unpacking „lessons learned”: Investigating failures and considering alternative solutions. *Knowledge Management Research & Practice*, 13(1), 17–30. <http://dx.doi.org/10.1057/kmrp.2013.26>
- Wiig, K. (1997). Knowledge management: an introduction and perspective. *Journal of Knowledge Management*, 1(1), 6–14. <https://doi.org/10.1108/13673279710800682>
- Williams, T. (2007). *Post-project reviews to gain effective lessons learned* (9th ed.). Project Management Institute.
- Williams, T. (2008). How do organisations learn lessons from projects – and do they? *IEEE Transactions in*

Engineering Management, 55(2), 248–266. <http://dx.doi.org/10.1109/TEM.2007.912920>

Wyrozębski, P. (2012). Doskonalenie procesów zarządzania projektami z wykorzystaniem narzędzi zarządzania wiedzą – stan obecny i perspektywy rozwoju. In Ł. Woźny (Ed.), *Ekonomia, finanse i zarządzanie w świetle nowych wyzwań gospodarczych* (pp. 209–235). Oficyna Wydawnicza SGH.

Wyrozębski, P. (2014). *Zarządzanie wiedzą projektową*. Difin.

Zedwitz, M. (2002). Organizational learning through post-project reviews in R&D. *R&D Management*, 32(3), 255–268. <http://dx.doi.org/10.1111/1467-9310.00258>

Zhang, L., Li, J., Li, A. Zhang, P., Nie, G., & Shi, Y. (2009). A new research field: Intelligent knowledge management. *International Conference on Business Intelligence and Financial Engineering*, 450–454. <http://dx.doi.org/10.1109/BIFE.2009.108>

Zhu, Y. (2020). Stock price prediction using the RNN model. *Journal of Physics: Conference Series*, 1650(3). <http://dx.doi.org/10.1088/1742-6596/1650/3/032103>

Zhu, T., Liao, Y., & Tao, Z. (2022). Predicting Google's Stock Price with LSTM Model. *Proceedings of Business and Economic Studies*, 5(5), 82–87. <http://dx.doi.org/10.26689/pbes.v5i5.4361>

Robert Pawlak holds a PhD in social sciences in the field of management and quality sciences, the SGH Warsaw School of Economics, and is an experienced project manager. His research interests revolve around knowledge management in organisations and learning from experiences in knowledge management, with a particular emphasis on machine learning methods.

Paweł Wyrozębski is a Professor at the SGH Warsaw School of Economics, a co-organiser and long-term lecturer in Postgraduate Studies in Project Management and Postgraduate Studies in Project Evaluation and Audit at the SGH Warsaw School of Economics. He specialises in project management, particularly in project management methodologies and standards, strategic project management and the operation of Project Management Offices (PMOs), project planning and control, IT support and tools for project management, and project organisation.

Ilona Pietras is a data engineer active in data processing design and implementation, cloud migration, and data modelling and visualisation. Her research interests pertain to a holistic approach to data-driven practices in enterprises.

Joanna Parys lectures at SWPS University in Warsaw and holds managerial roles at the consultancy firm EY Global Delivery Services. For years, she was affiliated with the banking and financial institutions sector, where she oversaw areas of project management, corporate governance, and organisational development. Her research interests centre on intellectual capital management in enterprises and on project management, with a specific emphasis on evaluating projects in terms of their profitability and effectiveness.

WE RECOMMEND



International Academic Conference on Teaching, Learning and E-learning, April 25–26 2025, Prague (Czech Republic) & online

International Academic Conferences are an important international gathering of scholars, educators and PhD students. Conference organized by the Czech Institute of Academic Education, z.s. in cooperation with the Czech Technical University in Prague.

Conference topics include: education, teaching, learning and e-learning education, teaching and learning, distance education, higher education, pedagogy, Erasmus and exchange experiences in universities, e-learning educational technology, educational games and software, and many others.

More information at: www.conferences-scientific.cz

"E-mentor" is one of the International Academic Conferences supporting journals.