Application of machine learning in cement price prediction through a web-based system

Adedeji Olushola Afolabi, Oluwanmayowa Abimbola
Department of Building Technology, College of Science and Technology, Covenant University, Ota, Nigeria

ABSTRACT
Cement is one of the most common building materials in the construction industry. Simultaneously, its price fluctuation can affect the success or failure of the construction project's performance. The study aimed to develop a web-based platform that uses machine learning algorithms on historical data of cement prices, petrol prices, diesel prices, interest rate, and exchange rate to predict future prices of cement products. The web-based learning platform was developed using hypertext markup language (HTML), cascading style sheet (CSS), MySQL, and hypertext preprocessor (PHP). For building a reliable machine learning model, python language was used to train the system. The front end, the back end, and the machine learning model were integrated with a flask python framework. A system block diagram was designed to show the web-based learning platform's interfaces. The web-based learning platform's system implementation led to the login page, the home page, database page, and cement price analytics interface. In training the machine learning model to make reliable cement price predictions, the study obtained an 80% fitted model in the linear regression. The web-based machine learning platform was able to predict the prices of cement. The rationale behind the machine learning prediction shown by the scatter plot diagram revealed that the cement increases by 250 naira biannually.

Keywords: Big data, Cement prices, Construction industry, Machine learning, Web-based system

1. INTRODUCTION
Accurate information serves as guidance to any pursuit. Organizations have cultivated the act of data mining to improve decision-making, planning, and the execution of projects and operations. While specific organizations have adopted knowledge management techniques, certain sectors are still lagging in embracing the process, especially in developing countries like Nigeria. With the advent of improved technology, which can capture large volumes of data, the term 'big data' is becoming the new rave. According to [1], big data in different sectors can increase its efficiency. The construction industry needs to realize that one of its best assets is its data. As Ismail et al. [2] pointed out, data can enhance the productivity of the construction industry. Analyzing these data in real-time can lead to smarter working, leading to operational performance in the sector [3]. Other sectors of the economy have increased their use of big data to enhance their performance. The study by Meulen [3] noted that companies had increased their investment in big data from 58-75% due to the vast potential it has to offer. Sectors such as the e-commerce space [4], education [5], manufacturing [6], public health [7], weather prediction [8], diabetes prediction (public health) [9],
earthquake prediction [10], and oil and gas [11] are participating in the big data advanced analytics to assist in its decision making and provide deep insight.

Specifically, there is a need to apply big data in planning and decision-making about building materials. Because of the critical role of building materials in their availability, volume needed high cost, quality, and crucial participation in erecting or constructing structure. Overall, construction materials’ costs account for up to 50 percent of the total construction sum [12], [13]. For individuals aspiring to own houses and the construction industry, the rising cost of building materials continuously poses a significant threat [14], [15]. Although, [16] argued that building materials’ price increase is a global phenomenon. From Sweden [17], the United States [18], South Africa [19], to Kenya [20], the cost of building materials continues to skyrocket. The study by Marzouk and Amin [21] stated that any changes in construction materials’ prices could significantly impact the cost of construction projects.

One such impact is the link established in [22], which showed that the rising cost of building materials is also responsible for time overrun experienced on most construction projects. Another impact is the high dispute rate between contractor and owner due to the rising cost of building materials [23]. In this study, we opined that with big data and machine learning algorithms, predicting the changes in construction materials prices and forecast actual values should be possible. Present studies on predicting construction material prices have been hinged on the notion put forward by [24] on predicting future construction materials’ prices. This is because construction materials’ prices are primarily volatile and subject to drastic changes that can occur quickly. Figure 1 showed the trend in cement prices from January 2004 to April 2020. Figure 1 shows that the cost of a 50 kg bag of cement had moved from 790 naira to 2600 naira, representing a 229 percent increase in the 16 years. Elfahham [25] stated that this economic uncertainty increases the potential financial risks and has a long-term impact on the construction business. This is one reason for fueling cost forecasting in the construction industry [26]. Considering the high capital investment required in construction projects and the number of years it takes to complete them, predicting future construction materials future prices is crucial [25]. Issa [24] opined that a contractor's ability to predict construction materials prices accurately would help them compete favorably in the construction industry and ensure the contractor's firm's success. The contractor's ability to predict construction material prices could help them during tendering and procurement planning stages in the construction delivery process [21]. Also, consultants could quickly inform construction project owners of upcoming projects' total costs with a prediction model [21]. Hwang et al. [26] observed that as building projects increase in size and the construction period's length is prolonged, there is a need to note construction material prices due to inflation factors.

This study aims to predict cement prices in the Nigerian construction industry through a machine learning-based approach. Specifically, a web-based learning platform was developed that uses machine learning algorithms on historical data in the form of knowledge management stored by a construction firm to predict future prices of cement products. The purpose is to help the construction firm predict cement prices for their upcoming projects in their bidding and procurement planning procedures. The approach used in this study is built on available historical data and numerical values of factors influencing cement prices. Similarly, Naser [27] used time-series data to predict oil prices in Iraq. This paper contributes to the existing scientific body of knowledge on automating cement price prediction using machine learning.

![Figure 1. Trends in cement prices (January 2004–April 2020)](image-url)

2. REVIEW OF RELATED LITERATURE

According to [28], [29], building materials used in the construction industry represent one of the most important factors affecting its successful performance. This makes the building construction sector a significant contributor to the construction industry's effectiveness primarily due to the associated cost in
procuring them [12], [30]. One of such building materials that are crucial to the construction industry is cement. The need for growth and development spurs the construction industry due to the continuous inflow of investment from different stakeholders it creates a high demand for cement products. This is Nigeria's case, a growing economy due to the new construction projects in housing and heavy civil construction that spring up daily. Ugochukwu et al. [30] noted that the demand for cement thrives in this kind of economies. Velumani and Nampoothiri [23] reported that the fluctuation in cement prices affects the construction industry's performance. Therefore, cement price prediction has taken center stage in the construction industry [23].

Several studies have pointed to the direction this paper aims to achieve to predict construction material prices. The study by Marzouk and Amin [21] used fuzzy logic and neural networks to predict the changes in building materials prices. Their system helps identify construction materials susceptible to changes that may affect contract price and expect future material cost changes. The primary data used were the changes in the material prices in the Egyptian construction market within a space of 10 years. The study by Issa [24] used artificial neural networks (ANN) to predict future construction material prices using the contractor's past costs of procured building materials and the conditions influencing these building materials. To forecast construction material prices, Hwang et al. [26] developed an automated time-series material cost forecasting (ATMF) system that uses the Box-Jenkins approach to produce a best-fitting model and forecast values for the construction material prices. The ATMF system developed by Hwang et al. [26] can help decision-makers in the construction sector predict future trends in construction material costs. Lee et al. [31] used a multivariate time series analysis to predict the prices of iron ore, a raw material in the production of steel products used in the construction industry. In predicting iron ore prices, system expected more than 2.3 times the past average values in previous studies. The study by Vinodhini and Velumani [32] performed an ANN test, SPSS analysis, and trend analysis on aluminum material prices' historical data. Their research showed that the use of ANN gave a more accurate price tag to aluminum materials. Specifically, Velumani and Nampoothiri [23] used three different methods in predicting the prices of cement in India. Their study used multiple regression methods, trend analysis, and ANN. Velumani and Nampoothiri [23] found that the use of ANN was more accurate in predicting cement prices in India. Even though most of the articles reviewed in this literature have used historical data on construction material prices and their influencing factors, none has used machine learning techniques in a web-based platform. This is one of the gaps this study intends to fill.

3. RESEARCH METHOD

Applying knowledge management and machine learning tools, this study seeks to develop an effective data collection process, data analysis, data processing, and predicting outcomes for improving construction operations. The experimental approach to this study used historical data of Nigeria's exchange rate to the dollar, bank interest rate, petrol prices, diesel prices, and cement prices from January 2004 to April 2020 in the web-based system that was developed. The historical data of exchange rate, bank interest rate, petrol prices, and diesel prices were obtained from the National Bureau of Statistics. In contrast, the cement prices were obtained from a private construction firm's knowledge management system as wholesale cement prices were purchased from January 2004 to April 2020. Figure 2 showed the trend of these explanatory variables used in this study. The computational model learning platform was developed within a web-based knowledge management system. The web-based system is divided into three major segments; the user interface, the back end, the database, the server, and the machine learning model.

In developing the web application, the user interface will be designed using hypertext markup language (HTML) and cascading style sheet (CSS). HTML is a web markup language applied to build the user interface framework. It can also be described as the foundation for front-end web development. HTML will design the login page, the home page, the analytic page, and the page to upload and download documents. CSS is an application that adds color and makes the web page presentable and enjoyable for the user. CSS works with the HTML tags and elements to be arranged appropriately, beautified, and unique to make the user experience excellent.

MySQL is a fast, easy-to-use relational database management systems (RDBMS) that many small and large companies use. MySQL is an inherently very efficient application. It manages a broad subset of the features of the most costly and efficient database packages. MySQL uses a simplified version of the popular SQL data language. MySQL works on many operating systems and many languages, including hypertext preprocessor (PHP), practical extraction and reporting language (PERL), C, C++, and JAVA. MySQL works very quickly and, even with substantial data sets, works perfectly. MySQL is very PHP-friendly, the most respected web development language. MySQL works very quickly and, even with large data sets, works fine. MySQL is very pleasant to PHP, the language most appreciated for web development. For building a reliable
machine learning model, python language will be used to train the system with the available data. The model will be a pickle file that serializes and de-serializes the model to deploy the website.

The integration of the three interfaces, the front end, the back end, and the machine learning model, will be integrated with a flask python framework. All HTML files, CSS files, pickle files, and python files will be compiled on the flask framework. Figure 3 showed the system block diagram of the web-based system for cement price prediction.

4. SYSTEM IMPLEMENTATION
This section dissected the newly built web-based system. It addressed the system's functional requirements, system specifications, system implementation, and training data in the developed web-based system. This section helps show the functional capacities, the design pattern and the interfaces of the web-based system for predicting cement prices.

4.1. Functional requirements
This section outlines the functional capacities of the web-based system for cement price prediction. The system aids in storing and retrieve construction information for construction firms. While keeping this file, they can build data (big data) to training the machine learning model to predict cement price. The web-based system majorly focused on cement price analytics. The web application is accessed from a laptop via a company-based URL for data collection, data retrieval, and cement price analysis. Making it web-based ensures it can be accessed from any location, is not susceptible to virus attacks, and can be monitored by multiple project stakeholders.
The user can access the web-based program with a username and password allocated by the system administrator via the login tab. After the user has successfully inserted the correct username and password, access is granted to the web-based system's tools. The following interface, after proper authorization, is the homepage. The homepage is characterized by a green header that contains the database link, cement price analytics, and the logout link.

The body of the database interface is divided into two. The upper half is to upload data into the server through the browse button and then the upload button, sending a success message after completing the file. The other half is to download files from the server. It is characterized by an input interface to input the file's name to download and a download button to successfully run that action.

The cement analytics interface is similarly characterized with a green header with two "home" links to redirect the website back to the home page while the log out to end the user's session. The interface body has a form to input the user's variables, which the computer will use to predict the price of cement for the year intended. The variables to be inputted on the form are the year of prediction, estimated exchange rate, estimated interest rate, the estimated price of petrol, and diesel price. After all, the variable has been properly inputted the predict button can be clicked. The variables will be sent to the machine learning model to analyze the variables. A result will be sent back on that interface stating the cement price of cement for the year's inputted year.

4.2. System requirements

The model-view-controller (MVC) design pattern was used to develop the web-based system for cement price prediction. HTML 5, CSS 3, and Python were the programming language used. Python was used because it is easy to learn, object-oriented. It works well for server programming and the machine learning model training, platform-independent, and moving from one computer system to another. The specification of the hardware was the operating system of Microsoft windows 10pro; version is 10.0.18362 Build 18362, system manufacturer is Dell Inc, a system model is latitude E7440, system Type is x64–based P.C., system SKU latitude E7440, the processor is intel® core ™ 13–4010u CPU @ 1.70 GHz, 1701 MHz, two cores, four logical processors installed RAM 8.00 GB. The system type is 64–bit operating system, x64-based processor, installed physical memory (RAM) is 4.00 GB, Total physical memory is 3.92 GB, available physical memory is 1.63 GB, total virtual memory is 4.61 GB, available virtual memory is 2.18 GB.

4.3. System implementation

This section showed the web-based system for cement price prediction interfaces and the interface summary. The web-based framework consists of four (4) key interfaces. The key interfaces are the login page, the homepage, the database page and the cement price analytics page.

4.3.1. Login interface

With the appropriate URL, the welcome page of the web-based system for cement price prediction should load. As the application is auth-based, the device's information and functionalities must be accessed by logging on. In the welcome tab, press the Login button. This takes the user to the login page. Sending a valid email address and password will take the user to the application's home page. If the user does not have system access, the user can contact the administrator. Making it internet-based ensures it is accessible from anywhere. The framework is not as vulnerable to virus attacks as mobile applications and can be controlled by multiple project stakeholders. Figure 4 showed the screenshot of the login page.

Welcome To C-KAM

Figure 4. Screenshot of the login interface
4.3.2. Home page interface

The home page offers a summary of the entire program and a point of reference to critical areas of the system. Figure 5 revealed the home page interface screenshot. The homepage presents a rundown of the new initiatives, a list of planned plans, and a dashboard to log out and access the database. The body of the home page has a button to redirect the user to cement price analytics.

![Home Page Interface](image)

Figure 5. Screenshot of the home page interface

4.3.3. Database interface

This interface is majorly concerned with two major activities in the database: to upload and download construction files that can be in various formats or extensions such as jpeg, pdf, txt, xlsx, csv, and png. This interface is critical for the user to store and retrieve vital information (knowledge). The functions of knowledge management system storage and sharing of information are essential parts of knowledge management. Figure 6 shows a screenshot of the database interface.

![Database Interface](image)

Figure 6. Screenshot of the database interface

4.3.4. Cement price analytics interface

The cement price analytics interface is the page that actively interacts with the machine learning model. The interface takes in the user's variables and sends this variable to the machine learning model to process and analyze. After the predict button clicks, the machine learning model's result is sent back to the cement price analytic interface. Figure 7 showed the screenshot of the cement price analytics interface. For instance, in this study, the cement prediction was tested in a web-based system to give a cement price prediction for 2024. Estimated values were inputted in the exchange rate variables at 450 naira to $1, the interest rate at 20%, petrol prices at 200 naira per liter, and diesel prices at 300 naira per liter. By pressing the prediction button, the machine learning platform gave the estimated cement price at 3,373.19 naira.

In the process of training the machine learning model to make reliable predictions. The dataset was adequately sorted and appropriately put to allow the system to use it to prepare the machine. The dataset was stored in an excel file called cement data.xlsx. Pandas, pickle, linear models from scikit-learn, and NumPy
had to be imported to train the machine learning model using the python software. The next step was to create a variable tag \( df \) that holds the dataset. The dataset features were split into \( x \) and \( y \). \( x \) represents the independent features while \( y \) is the dependent feature (the variable predicted). On the next line of code, \textit{linear_model} was used. Linear Regression helped to create the relationship between feature \( x \) and \( y \). After which, this relationship will become a prediction formula. The model was stored in a pickle format before imported to the other line of code for predictions. The line of code \texttt{pickle.dump(regr,open(cement_model.plk, 'wb'))} as shown in Figure 8 was used to convert the model to a pickle format. By using \texttt{regr.fit(x,y)}, it helped determine how fitted the model in which the study obtained 0.8. The machine learning model's accuracy training at 80% shows that the model is an excellent model for predicting cement prices. Previous studies that have used linear regression in their machine learning can be found in [33].

![Figure 7. Screenshot of the cement price analytics interface](image)

![Figure 8. Screenshot of the cement price analytics machine learning model training](image)

\begin{verbatim}
In [20]:
import pandas
import pickle
from sklearn import linear_model
import numpy
import matplotlib.pyplot as plt

numpy.random.seed(2)

df = pandas.read_excel("cement data.xlsx")
X=df[['year','exchange rate','Interest rate','Petrol','Diesel']] y=df['Price']

regr = linear_model.LinearRegression()
regr.fit(X,y)
pickle.dump(regr, open('cement_model.plk', 'wb'))
\end{verbatim}
Furthermore, Figure 9 graphically shows the relationship between petrol and cement price. After the model training with cement price data from January 2004 to April 2020, the scatter plot diagram shows the rationale behind the machine learning predictions. The y-axis represents the price of cement while the x-axis represents the price of petrol. On the y-axis, the cost of cement spaces 250 naira per gridline, while on the x-axis, the petrol price is spaced 20 naira per gridline from the scatter plot.

![Petrol vs Cement Price](image)

**Figure 9. Screenshot of the cement price analytics machine learning model training vs. petrol price**

Figure 10 graphically shows the relationship between the interest rate and the price of cement. After the model training with cement price data from January 2004 to April 2020, the scatter plot diagram shows the rationale behind the machine learning predictions. The y-axis represents the price of cement, while the x-axis represents the interest rate. On the y-axis, the cost of cement spaces 250 naira per gridline, while on the x-axis, the interest rate is spaced 1% per gridline from the scatter plot.

![Interest Rate vs Cement Price](image)

**Figure 10. Screenshot of the cement price analytics machine learning model training vs. interest rate**
Figure 11 graphically shows the relationship between diesel price and the price of cement. After the model training with data on cement price from January 2004 to April 2020, the scatter plot diagram shows the rationale behind the machine learning predictions. The y-axis represents the price of cement, while the x-axis represents the price of diesel. On the y-axis, the cost of cement spaces 250 naira per gridline, while on the x-axis, diesel's price is spaced 50 naira per gridline from the scatter plot.

Figure 11. Screenshot of the cement price analytics machine learning model training vs. Diesel prices

Figure 12 graphically shows the relationship between the exchange rate and the price of cement. After the model training with cement price data from January 2004 to April 2020, the scatter plot diagram shows the rationale behind the machine learning predictions. The y-axis represents the price of cement, while the x-axis represents the exchange rate. On the y-axis, the cost of cement spaces 250 naira per gridline, while on the x-axis, the exchange rate is spaced 25 naira per gridline from the scatter plot.

Figure 12. Screenshot of the cement price analytics machine learning model training vs. exchange rate
Figure 13 graphically shows the relationship between the year and the price of cement. After the model training with cement price data from January 2004 to April 2020, the scatter plot diagram shows the rationale behind the machine learning predictions. The y-axis represents the price of cement, while the x-axis represents the year. On the y-axis, the cost of cement spaces 250 naira per gridline, while on the x-axis, the year is spaced biannually per gridline from the scatter plot.

```python
pickle.dump(regr, open('cemnt_model.plk', 'wb'))
plt.scatter(df['year'], y, color = "red")
plt.title("Year Vs Price of Cement", fontsize=14)
plt.xlabel("Year", fontsize=14)
plt.ylabel("Price of Cement", fontsize=14)
plt.grid(True)
plt.show
```

Out[20]: <function matplotlib.pyplot.show(*args, **kw)>

Figure 13. Screenshot of the cement price analytics machine learning model training with the scatter diagram (year vs. price of cement)

5. CONCLUSION

The study aimed to develop a web-based learning platform that uses machine learning algorithms on historical data of cement prices, petrol prices, diesel prices, interest rate, and exchange rate to predict future prices of cement products. The web-based learning platform was developed using HTML, CSS, MySQL, and PHP. For building a reliable machine learning model, python language was used to train the system with the available data. The front-end, the backend, and the machine learning model were integrated with a flask python framework. A system block diagram was designed to show the web-based learning platform's interfaces for predicting cement prices. The web-based learning platform's system implementation led to the login page, the home page, database page, and cement price analytics interface. In training the machine learning model to make reliable cement price predictions, the study obtained an 80% fitted model in the linear regression. The rationale behind the machine learning prediction shown by the scatter plot diagram revealed that the cement increases by 250 naira biannually. Besides, when petrol prices are increased by 20 naira, the interest rate increases by 1%, diesel prices are increased by 50 naira, and the exchange rate is increased by 25 naira, cement prices increase 250 naira. The study recommends the increased use of machine learning in predicting building materials prices to aid in bid preparation, construction estimates, and construction planning. For future research, other web application systems can be developed to consider other building materials, such as creating a prediction model for steel reinforcement prices, sandcrete blocks, granite, sand, and so on.

ACKNOWLEDGEMENTS

The authors appreciate Covenant University through the Centre for Research, Innovation, and Discovery (CU-CRID) in making this submission Open Access by contributing to the article processing charge.
REFERENCES


Application of machine learning in cement price prediction

Adedeji Olushola Afolabi received the B.Sc. degree in Building from University of Lagos, Nigeria, in 2009, M.Sc. in Construction Management from University of Lagos, Nigeria, in 2012, and Ph.D. Building Technology (Construction Management) from Covenant University, Nigeria in 2018. Currently, he is a Senior Lecturer at the Department of Building Technology, Covenant University. His research interests include digital construction, gender studies and sustainable construction. He can be contacted at email: adedeji.afolabi@covenantuniversity.edu.ng.

Oluwamayowa Abimbola received the B.Sc. degree in Building Technology from Covenant University, Nigeria. He obtained a M.Sc. degree in Building Technology (Construction Management) from Covenant University, Nigeria in 2020. Currently, he is practicing in the Nigeria Construction Industry. Also, he designs offline and web-based systems used in the construction industry. He can be contacted at email: mabimbola91@gmail.com.