






Correlation Between Big Data and Cloud in the Transactions Environment

Anassin Chiatsè Mireille Patricia*^{}, Saha Kouassi Bernard^{}, Kra Lagasane^{}

Department of Science and Technology, College of Computer, Université Alassane Ouattara de Bouaké,
Bouaké 01 01 BP V18, Côte d'Ivoire

Corresponding Author Email: anassin.patricia@uao.edu.ci

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290308>

ABSTRACT

Received: 10 April 2024
Revised: 27 May 2024
Accepted: 11 June 2024
Available online: 20 June 2024

Keywords:

big data, cloud computing, security, machine learning (ML), Parameterized Correlation Approach (PCA), transactions environment

Technological advances are opening up opportunities for megadata processing. Cloud computing is a catalyst for significant cost reduction. What's more, it's a technology available to companies of all sizes. With the emergence of big data, there are sophisticated analysis possibilities that influence both processes and program operation. This article presents a new cloud computing infrastructure for big data analysis in the transactional environment. This Parameterized Correlation Approach (PCA) will enable real-time process security and optimization. The study used historical transactional data and machine learning optimization models to develop and deploy our architecture. The results of the experiment showed that the machine learning-based algorithm promotes excellent transactional prediction capabilities.

1. INTRODUCTION

Computer industry have considerably transformed the way companies operate, particularly in the area of data management. Big data is an essential element shaping the transactional environment of organizations [1]. Indeed, the vast amount of information they generate requires complex analysis tools. They are artificial intelligence and machine learning [2]. Over recent years, these pairs have significantly revolutionized the industrial world. In this case, they play a role in the more efficient and cost-effective collection and processing of voluminous data for large-scale cloud computing.

Machine learning is at the heart of this research, offering robust predictive models using a variety of algorithms [3, 4]. The efficiency and objectivity of these techniques often surpass those of traditional tools, making corporate prevention and decision-making more accurate and effective [5, 6].

In addition, the Big Data revolution has produced characteristics in terms of volume, variety and velocity, providing a wealth of information for ML algorithms [7, 8]. The rise of the Internet of Things (IoT) [9], Such applications have become essential to organizational processes. It is now possible, at relatively low cost, to transfer huge quantities of data to the cloud for storage and analysis, thanks to robust and accessible infrastructures.

To optimize their operations and exploit the full potential of their assets, companies need to know the correlation between big data and the cloud. By seamlessly integrating big data analytics with cloud storage [10], companies can gain valuable insights, improve decision-making processes and increase overall efficiency.

However, Businesses face a number of challenges in

integrating and optimizing transactional processes [11-13]. Key issues include data security, real-time optimization, and the ability to process and analyze massive volumes of data efficiently. It is therefore crucial to develop a cloud computing architecture capable of meeting these challenges, using machine learning algorithms to analyze historical transactional and optimize processes in real time.

In this study, we will examine the correlation between big data and clouds in the transactional environment, exploring how this synergy can drive innovation and competitive advantage in modern business operations. This Parameterized Correlation Approach (PCA involves setting up a new transactional analysis infrastructure. It will also enable processes to be secured and optimized in real time, using machine learning algorithms on historical data.

In other words, the proposed architecture aims to combine legacy-based process analysis, which is carried out mostly offline, or in time samples, with multiple sensory data streams describing storage and retrieval process states. Data must be robust to infrequent updates. Inconsistent measurements are designed for highly variable intervals. The model is also capable of handling correlation between different processes.

The article is organized as follows: section 2 presents the state of the art, section 3 entitled new cloud computing architecture for Big Data in the transactional environment while highlighting the proposed model. Section 4 presents the various results obtained. Finally, we draw the general conclusions of our study.

2. RELATED WORKS

The integration of big data and cloud into the transactional

environment has revolutionized modern business operations. Through in-depth data analysis using tools such as Elastic search and Kibana [14], business trends are based on consumer behavior, market variations and product performance. This comprehensive analysis enables organizations to make informed decisions that boost sales and improve customer satisfaction.

This is where customer relationship management (CRM) comes in, and is an essential research topic. It has made companies compete more effectively [15, 16]. Researchers first proposed the concept of perceived value [17]. Customers appropriate this entity when they buy products. In the end, services obtain an overall assessment of the costs paid. At present, there are two ways of defining this value. Firstly, at purchasing level [18].

This value is the difference between quantity and total cost. The other is based on a business perspective [19]. The (CLV) is determined over the life cycle. It refers to all the benefits created by customers for the company throughout the process of maintaining their relationship with the company [20]. In 1985, Petit et al. [21] first proposed a definition based on the traditional method during the period of future consumption. They define the measurement model for studying customer value from a business perspective.

However, the researchers noted the shortcomings of the (CLV) model in the following aspects: (1) all models are too conceptual and idealized. They cannot be applied in practice. In particular, some are more influenced by other factors and difficult to calculate [22]. (2) Considers the benefits brought to the company. (3) Sales production channels are inaccessible. Customers will have more choice. As a result, competition between the various parts intensifies. What's more, developing a complete life cycle requires available, up-to-date data.

In addition, the adoption of deep learning technologies has further amplified impact of big data on transactional processes [23]. By leveraging the capabilities of platforms, companies can access scalable resources for data storage, processing and machine learning, enabling improved efficiency, predictive analytics and strategic competitiveness. In this way, various algorithms are used to predict customer behavior and transaction value.

The authors [24] built a prediction model based on machine learning. They compared logistic regression and decision trees, to select the best features in the case of telecommunications. As for the study of Charalambidis et al. [25], they compared the performance of different methods to solve this problem. Experimental results show that SVM (Support Vector Machine) is superior to artificial neural networks, decision trees and Bayesian classifiers in terms of coverage rate. Similarly, Cheng et al. [26] used these approaches to predict future customer purchases, proving that feature extraction is very important.

These various works [26] have revealed that the machine learning-based algorithm has excellent predictive capabilities of customer behavior than traditional interpretive models.

Furthermore, the seamless integration of these workflows not only transforms operational workflows, but also improves decision-making and customer engagement, consolidating its central role in shaping the transactional landscape of the digital age.

What's more, their use in transactions offers numerous benefits that improve business operations [27]. Firstly, it

enables organizations to quickly and efficiently improve important features, providing valuable insights into valuable information on customer behavior, market trends and operational efficiency. This approach [28] enables companies to make informed decisions in real time, thus satisfying strategic planning.

Scalable, cost-effective storage solutions are also available, eliminating the need for investment in infrastructure [29]. Businesses can easily access data, making transactional processes more flexible and mobile. Combined, companies can streamline transactions, improve customer experience and drive innovation in the competitive marketplace [30]. Transactional synergy is a compelling argument.

This article, based on the above-mentioned research, takes the view that, firstly, traditional methods are still relevant. The prediction effect of different approaches is biased by considering preprocessing; secondly, for different data sets, the performance of various algorithms is different, and the prediction of a single method shows some discrepancies. Therefore, combination is an ideal modeling strategy. With this in mind, we focus on similarity approaches in the general area of cloud-based security processes and prediction parameters for information provided to end-users.

3. METHODOLOGY

The presentation, on the one hand, of a modeling architecture based on cloud review and, on the other, of an optimal process control distribution configuration. The architecture plays the role of managers in modern networks. Local computers can access and execute machine learning algorithms while synchronizing them. Existing steps remove time limitations due to complexity. Figure 1 illustrates the overall system architecture.

3.1 Method architecture

The storage and integration of Big Data makes it possible to visualize and reuse information for analytical purposes. Additionally, this data is partitioned for machine learning reasoning. Furthermore, this information leads to the discovery of irregularities such as anomalies and advanced redundancies during the transfer to the cloud. For this reason, a security system should be considered which allows the information received to be evaluated. First, before sending the data to the cloud, the security system analyzes and sorts all the data received. In addition, it supports data authenticity and confidentiality to protect user privacy. Then the system will produce a summary and index the different anomalies to process them. The processing is done using the aggregation calculation and the interaction parameters (see the mathematical model). In addition, in the event of fires or disconnections, assistance in performance management is established and made available to customers. Finally, the information is used by end users during their various requests in real time.

Figure 1 summarizes everything written above. That is to say the data is first stored in Big Data, and then these are processed by the security system set up and finally sent to the Cloud to be deployed.

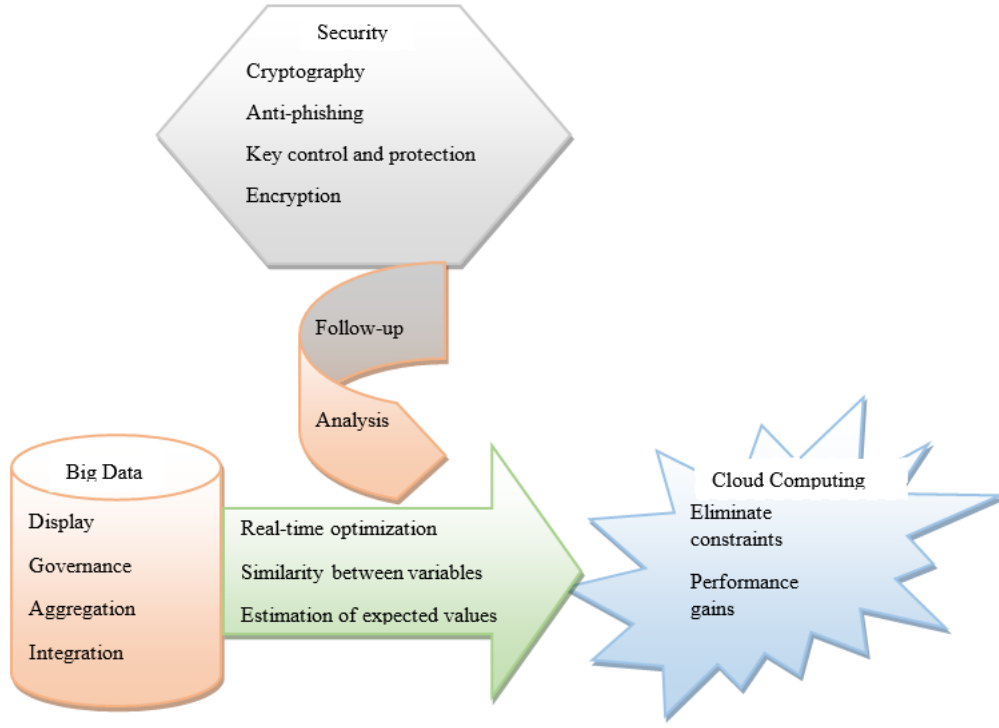


Figure 1. The steps for transferring data to the cloud

3.2 Mathematical formulation of the method

The study of the correlation between cloud security parameters and big data quality standardized PCA (Parameterized Correlation Approach) is measured by the tendency of end users to have the traceability and response of their queries over time. The number of transactions made over a period of time is calculated according to the probability of similarity coverage of the variables. We base ourselves on two hypotheses [2, 8]:

If the information demand is adjusted for partitioning operations, then the function is constant throughout time.

Otherwise, if the number of transactions is lower or higher than the partitioning operations, the similarity parameters and variables have a low or high rate compared to the defined threshold.

The information provided about partitioning operations is represented by the characteristics of the following function:

$$S_{l,k} = \sum_{k=0}^{D_t} \binom{D_t}{k} J^k (1-J)^{D_{x,t}-k} \left| \frac{k}{D_t} - J_t \right| \quad (1)$$

$S_{l,k}$ is the function for adjusting the different parameters. Its value is zero if D_t does not exist. It follows an evolution and depends on the characteristics of D_t .

D_t determines partition operations when transferring data to the cloud

J the number of transactions during queries

k determines the intensity of the parameter distribution on partition operations

To predict customer behavior, the expected number of transactions over a period is given by the following formula:

$$E(Y(t)|X = x, t, T, \gamma, v, i, \theta) = \frac{i + \theta + x - 1}{i - 1} \times \left[1 - \left(\frac{v + T}{v + T + t} \right)^{\gamma+x} F_1 \left(\gamma + x; \theta + x; i + \theta + x - 1; \frac{t}{v + T + t} \right) \right] / 1 + \sigma_{(x>0)} \frac{q}{\theta + x - 1} \left(\frac{v + T}{v + t_x} \right)^{\gamma+x} \quad (2)$$

E refers to the value expected by the customer during each transaction

X highlights the expected conditional number of transactions and indicates the frequency of requests

t_x information update time and availability

T the time elapsed between different updates during Big Data and cloud transfer

γ, v transaction rate of different clients due to gamma distribution

i, θ anomaly rate from the beta distribution

This mathematical formulation is used in various cases:

Case 1: Transaction prediction

The mathematical formulation is used to predict the number of transactions expected over a given period. Analysts can estimate the frequency of transactions based on historical data and current conditions. In practice, this model can be applied in financial transaction management systems to improve forecast accuracy and optimize transaction processes.

```
def preprocess_data (data):
```

```
    cleaned_data = clean (data)
```

```
    normalized_data = normalize (cleaned_data)
```

```
    return normalized_data
```

```
def model_transactions (data,gamma_v,theta):
```

```
    X = calculate_conditional_expected_transactions (data)
```

```
T = calculate_time_intervals (data)
model = gamma_v*theta*(X + T)
return model
```

Case 2: Adjusting cloud security parameters

The parameter adjustment function is used to optimize data partitioning operations when transferring data to the cloud. In fact, the following function adjusts the parameters according to the demand for information and the number of transactions. This formula is used to ensure that operations are optimized, improving the security and efficiency of data transfer to the cloud. For example, a financial services company could use this function to secure customer transactions while minimizing security risks.

```
def adjust_security_params (D_t, J, k):
    if D_t != 0:
        S_lk = (D_t*J*k*(1-J)*D_t-k*k*D_t-J*T) / (D_t-J*T)
    else:
        S_lk = 0
    return S_lk
def predict_behavior (model, gamma, theta):
    expected_transactions = model*(gamma + theta)
    return expected_transactions
```

Case 3: Customer data analysis

She is also applied to analyze customer behavior and predict their value over time. Using variables such as transaction rates and detected anomalies, companies can segment their customers and target more effective marketing campaigns. For example, a telecoms company could use this approach to predict which customers are likely to switch providers, and then proactively intervene to retain them.

```
def optimize_partitioning (data):
    optimized_data = partition_data (data)
    secure_data = secure(optimized_data)
    return secure_data
```

This mathematical and algorithmic approach ensures efficient and secure data management in a cloud environment, while optimizing performance and forecast accuracy.

In the next section, we'll look at how she is being applied in practical contexts to improve prediction accuracy and process efficiency in big data and cloud computing.

4. EXPERIMENTS AND RESULTS

These developments were carried out in close connection with the process industry. Similar techniques with a similar working climate are considered to survey the proposed calculation. The security system processes information at the big data level before sending it to the cloud. The next methodology is the message-based exception identification model. Then it cleans up the information as it moves without laying out the model in advance.

4.1 Evaluation of approaches according to parameters

In this part, the evaluation of the proposed technique is highlighted. The data used comes from an internet platform for online insurance sales (<https://archive.ics.uci.edu/ml/datasets/Online+Retail+II>).

For our various simulations, we used Python software. It was performed on a processor that met the following requirements: CPU: 3.00 GHz Intel (R) Core (TM) i5-8500, the relevant criteria of the simulation are listed in the table The values of the parameter thresholds are used as independent and

disjunctive to predict the response of the target variable. The relevant simulation criteria are listed in the various tables.

Table 1 presents the security parameters to evaluate the performance of the two algorithms. These algorithms are used in the transactional environment to analyze the predictive effect of information provided to customers. In the results obtained, the precision and recall are close to 1. This indicates that the PCA makes it possible to maximize the cost of producing the final usage information compared to SVM (Support Vector Machine). As for accuracy and F-Measure, the results are biased. This is explained by the heterogeneity of Big Data. In this case the clients minimize their effects for the different requests. In addition, the accessibility time is much reduced.

Table 1. Evaluation of the two approaches according to security parameters

Approaches	Exactness	Precision	Reminder	F-Measure
SVM	0.955	0.970	0.980	0.975
PCA	0.987	0.970	0.989	0.985

The results obtained show a comparison between the two methods of evaluating online transactions.

Comparative analysis of different parameters

Accuracy: This measures the percentage of correct predictions among all predictions made. It is an overall measure of classification performance.

SVM: With an accuracy of 0.955, demonstrating a good ability to classify transactions correctly.

PCA: The PCA algorithm outperforms SVM with an accuracy of 0.987, indicating better overall performance in terms of correct classification.

Precision: This indicates the proportion of correct positive predictions in relation to the total number of positive predictions. High precision means fewer false positives.

SVM: Precision of 0.970 suggests that the majority of transactions predicted as positive are indeed correct.

PCA: With an accuracy of 0.990, it shows an even greater ability to correctly identify positive transactions, thus reducing false positives.

Recall: This measures the proportion of true positives correctly identified among all items that are actually positive. High recall means fewer false negatives.

SVM: Recall of 0.980 indicates that it is able to capture most positive transactions.

PCA: PCA achieves a recall of 0.989, showing a slight improvement over SVM in the detection of positive transactions, thus reducing false negatives.

F-Measure: is the harmonic mean of precision and recall. It combines the two measures into one, to assess the balance between precision and recall.

SVM: With an F-Measure of 0.975, SVM maintains a balance between precision and recall.

PCA: PCA's F-Measure of 0.985 confirms its superiority in terms of precision/recall balance, consolidating its high performance on all parameters evaluated.

Ultimately, the PCA algorithm outperforms the SVM algorithm in all the safety parameters evaluated, namely accuracy, precision, recall and F-Measure. This superiority can be attributed to PCA's ability to better handle the heterogeneity of Big Data, enabling a significant reduction in classification errors. These results suggest that PCA is a robust and effective method for analyzing online transactions,

particularly in complex environments where predictive accuracy and reliability are crucial.

PCA's performance in terms of reducing false positives and negatives, as well as its ability to maintain high accuracy and recall, make it a preferred choice for transaction security applications in Big Data and Cloud Computing environments.

Table 2 lists the performance of the different parameters. These show a similarity closer to 1. These results are obtained from the BG/NBD (Geometric Beta/Negative Binomial Distribution) model following the values of the different distributions.

Table 2. Evaluation of the two approaches according to the values of the distributions

Approaches	γ	ν	i	θ
SVM	9.964	1.887	0.071	1.293
PCA	9.976	2.031	0.094	1.567

Statistical analysis based on distribution

γ (Gamma): Transaction rate of different customers according to gamma distribution. It indicates the frequency of transactions.

SVM: The gamma parameter for SVM is 9964, indicating the transaction rate of different customers according to the gamma distribution.

PCA: PCA has a value slightly higher than 9976, suggesting better management of customer transaction rates.

ν (Value): Measure of transaction variability. It represents the dispersion or heterogeneity of transactions.

SVM: The SVM value of 1887 shows the model's performance in terms of transaction variability.

PCA: With a value of 2031, PCA shows higher variability, which may reflect better adaptability to transaction variations.

i (Anomaly): The anomaly rate derived from the beta distribution, indicates the frequency of abnormal transactions detected.

SVM: The anomaly rate derived from the beta distribution for SVM is 0.0071, indicating a low level of detected anomalies.

PCA: PCA, with a value of 0.0094, detects slightly more anomalies than SVM, which may signify a better sensitivity to anomalies.

θ (Theta): Measure of anomaly correction. Represents then the algorithm's ability to correct or handle detected anomalies.

SVM: SVM's theta parameter is 1293, representing a certain level of performance in terms of anomaly correction.

PCA: PCA, with a theta of 1567, shows an improved ability to correct anomalies compared to SVM.

We find that the analysis of safety parameters using the BG/NBD model shows that the PCA approach outperforms the SVM approach in almost all aspects evaluated. PCA demonstrates better transaction rate management, greater variability, and better anomaly detection and correction. These results suggest that PCA is a more robust and reliable method for processing transactional data in the context of Big Data and Cloud Computing.

Table 3 shows the character of the two algorithms. The degree of confidence is approximately equal for the two algorithms. In addition, the deviation error between the different parameters is very low. This explains that the two algorithms are better at managing the quantity of information contained in the Cloud. Both algorithms are error tolerant in the case where the prediction parameters are very high as well as the indexing power.

Table 3. Evaluation of the two approaches according to the correlation of the different parameters

Approaches	Deviation Error	Indexing Power	Anomaly Report	Degree of Confidence
SVM	0.025	0.738	0.315	0.991
PCA	0.015	0.801	0.143	0.996

Method of evaluating the various parameters

Deviation error: Measures the difference between predicted and actual values. A low deviation error indicates high prediction accuracy.

SVM: The deviation error for SVM is 0.025, indicating a moderate level of deviation from predictive values.

PCA: PCA performs better with a lower deviation error of 0.015, suggesting higher prediction accuracy.

Indexing power: This indicates the algorithm's ability to index and organize data efficiently. A higher value means better data structuring.

SVM: its indexing power is 0.738, reflecting its ability to index and organize data.

PCA: With a value of 0.801, PCA demonstrates superior indexing power, which means better data management and structuring.

Anomaly ratio: This indicates the frequency of anomalies detected in relation to the total number of transactions. A lower anomaly ratio indicates better anomaly management.

SVM: The anomaly ratio for SVM is 0.315, indicating the frequency of detected anomalies.

PCA: PCA shows a lower anomaly ratio of 0.143, indicating better performance in anomaly detection and management.

Confidence level: A measure of the reliability of the predictions made by the algorithm. A higher degree of confidence indicates more reliable and accurate predictions.

SVM: The confidence level for SVM is 0.991, indicating high prediction reliability.

PCA: PCA achieves a slightly higher confidence level of 0.996, indicating very high reliability and accuracy.

Consequently, the analysis of correlation parameters shows that the PCA approach outperforms the SVM approach in almost all aspects evaluated. PCA demonstrates better accuracy with lower deviation error, greater indexing power, reduced anomaly ratio and slightly higher confidence. These results suggest that PCA is more effective at managing and analyzing transactional data, offering a more robust solution for integrating Big Data and Cloud Computing.

In summary, the comparative analysis of PCA (Parameterized Correlation Approach) and SVM (Support Vector Machine) methods is carried out using the results of three tables, each evaluating different aspects of the performance of the two approaches. The parameters defined in each table enable a comprehensive and detailed evaluation of the performance of PCA and SVM methods. They cover various essential aspects, such as prediction accuracy, anomaly handling, and the efficiency of indexing and data structuring. These measures make it possible to compare and determine the relative effectiveness of the two methods in the context of Big Data and Cloud Computing.

4.2 Exploratory discovery of different approaches

In this simulation, we set the strange information level to 80% for each partition operation. Exploratory findings show that the impacts of information volume on time delay are insignificant. Its effect on limiting time and storage space

consumption is recognizable when the amount of information is huge. The higher the amount of information, the better the impact of the proposed model adjustment, which is why this strategy can reduce the delay and consumption.

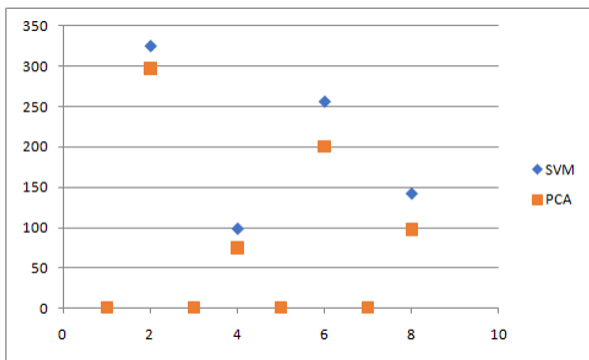


Figure 2. Observation of the parameters in Table 1 on the SVM and PCA approaches

Figure 2 presents the parameterized loading and processing curves of the two algorithms. For the different tests, the execution time of the PCA (Parameterized Correlation Approach) algorithm is closer to a few seconds. That is to say that the information transmitted is processed to definitively put an end to the anomalies. Furthermore, the prediction of anomalies is very negligible. On the other hand, the SVM reflects the prediction effect while maximizing the cost of anomalies.

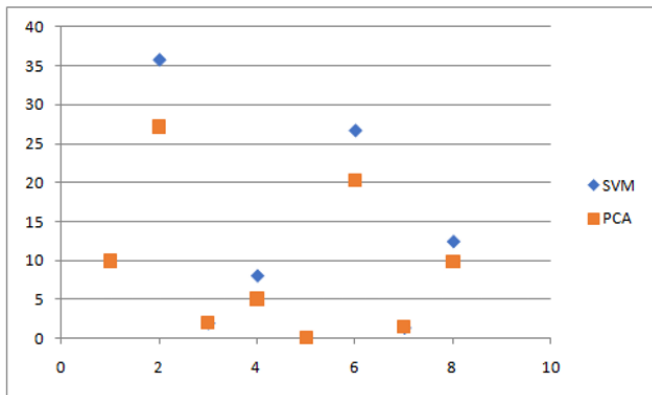


Figure 3. Observation of the values in Table 2 on the SVM and PCA approaches

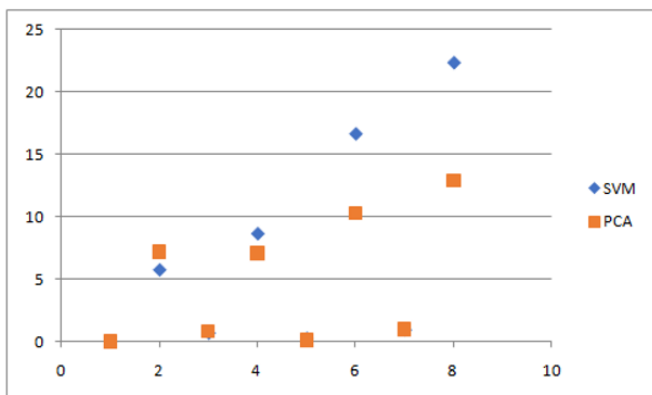


Figure 4. Observation of the correlation of the parameters in Table 3 on the SVM and PCA approaches

Figure 3 shows the prediction of the number of transactions over the next twenty-four hours based on the actual parameter values. It highlights the trend and visualization of the matrix.

Figure 4 shows the information available on the cloud. This is sufficiently processed and made available to customers in complete security. In addition, this explains the position of the SVM point clouds which increases in a linear and significant manner. By contrast, those of PCA gradually decrease.

5. CONCLUSION

In this paper, new cloud architecture for optimizing big data is proposed in the context of transactional data integration and processing. This Parameterized Correlation Approach (PCA) solved the problem of predictive efficiency in exploring customer queries in real time. We used the BG/BND model and machine learning algorithm for the functional analysis of the process and the implemented security system. Compared with those of the SVM algorithm, our approach shows a significant reduction in time and storage space consumption when the amount of data is large. The approach can be improved by using other data security settings. Finally, our method can be extended to other domains such as agriculture while evaluating the parameter conditions of this sector.

ACKNOWLEDGMENT

The authors gratefully acknowledge ALASSANE OUATTARA University. This research is a project whose funding validation is in progress. We also gratefully acknowledge the directors of the various laboratories for promoting the research.

REFERENCES

- [1] Zhang, X., He, Y., Pan, L., Yao, Z. (2022). Sales data analysis of cloud computing products based on big data. *IFAC-PapersOnLine*, 55(10): 1404-1409. <https://doi.org/10.1016/j.ifacol.2022.09.587>
- [2] Priya, P.S., Malik, P., Mehbodniya, A., Chaudhary, V., Sharma, A., Ray, S. (2022). The relationship between cloud computing and deep learning towards organizational commitment. In *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Gautam Buddha Nagar, India, pp. 21-26. <https://doi.org/10.1109/iciptm54933.2022.9754046>
- [3] Zhang K, Lu, Q., Chen, Yang, Z., Chen, Y., Li, H., Yin, X. (2023). Modality analysis and algorithm design of stator short-circuit fault set for large compressed air energy storage generators. *Energy Reports*, 9: 58-65. <https://doi.org/10.1016/j.egy.2022.10.357>
- [4] Li, Z., Chen, X., Shi, S., Zhang, H., Wang, X., Chen, H., Li, L. (2022). DeepBSA: A deep-learning algorithm improves bulked segregant analysis for dissecting complex traits. *Molecular Plant*, 15(9): 1418-1427. <https://doi.org/10.1016/j.molp.2022.08.004>
- [5] Huang, S., Ali, N.A.M., Shaari, N., Noor, M.S.M. (2022). Multi-scene design analysis of integrated energy system based on feature extraction algorithm. *Energy Reports*, 8(S6): 466-476.

- <https://doi.org/10.1016/j.egyr.2022.03.161>
- [6] Kabir, M.F., Chen, T., Ludwig, S.A. (2023). A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. *Healthcare Analytics*, 3: 100125. <https://doi.org/10.1016/j.health.2022.100125>
- [7] Sun, Y., Liu, H., Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon*, 9: e13384. <https://doi.org/10.1016/j.heliyon.2023.e13384>
- [8] Ravindhar, N.V., Sasikumar, S. (2022). An effective monitoring, storage and analyze on industrial process on cloud big data by data publishing in industrial wireless sensor network. *Measurement: Sensors*, 24: 100525. <https://doi.org/10.1016/j.measen.2022.100525>
- [9] Qu, C., Luo, W., Zeng, Z., Lin, X., Gong, X., Wang, X., Li, Y. (2022). The predictive effect of different machine learning algorithms for pressure injuries in hospitalized patients: A network meta-analysis. *Heliyon*, 8: e11361. <https://doi.org/10.1016/j.heliyon.2022.e11361>
- [10] Wu, B., Tian, F., Zhang, M., Zeng, H., Zeng, Y. (2020). Cloud services with big data provide a solution for monitoring and tracking sustainable development goals. *Geography and Sustainability*, 1(1): 25-32. <https://doi.org/10.1016/j.geosus.2020.03.006>
- [11] Goldin, E., Feldman, D., Georgoulas, G., Castaño, M., Nikolakopoulos, G. (2017). Cloud computing for big data analytics in the Process Control Industry. In 2017 25th Mediterranean Conference on Control and Automation (MED), Valletta, Malta, pp. 1373-1378. <https://doi.org/10.3390/app122010567>
- [12] Jarosz, J. (2022). Big data and cloud computing: roles and relationships, techniques and tools. *Journal of Data Analytics*, 1(1): 33-41. <http://doi.org/10.59615/jda.1.1.33>
- [13] Zhao, X. (2022). Research on management informatization construction of electric power enterprise based on big data technology. *Energy Reports*, 8(S7): 535-545. <https://doi.org/10.1016/j.egyr.2022.05.124>
- [14] Henden, H., Irawatia, R., Indra, I., Dewi, D.A., Kurniawan, T.B. (2023). Big data analysis using elasticsearch and Kibana: A rating correlation to sustainable sales of electronic goods. *HighTech and Innovation Journal*, 4(3): 583-591. <http://doi.org/10.28991/HIJ-2023-04-03-09>
- [15] Hameurlain, A., Küng, J., Wagner, R., Sakr, S., Wang, L., Zomaya, A. (2015). *Transactions on Large-Scale Data-and Knowledge-Centered Systems XX: Special Issue on Advanced Techniques for Big Data Management*. Springer. <https://doi.org/10.1007/978-3-662-46703-9>
- [16] Xu, N., Qin, R., Song, S. (2023). Point cloud registration for LiDAR and photogrammetric data: A critical synthesis and performance analysis on classic and deep learning algorithms. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 8: 100032. <https://doi.org/10.1016/j.ophoto.2023.100032>
- [17] Wu, B.G., Yang, L., Chen, Y.B. (2020). An empirical study of purchase rate and dropout rate between mobile and PC customers. *Journal of Systems & Management*, 29(5): 924-933. <https://doi.org/10.3969/j.issn.1005-2542.2020.05.010>
- [18] Wang, Y., Xia, Y., Fang, Q., Xu, X. (2018). AQP++: A hybrid approximate query processing framework for generalized aggregation queries. *Journal of Computational Science*, 26: 419-431. <https://doi.org/10.1016/j.jocs.2017.05.001>
- [19] Mavruk, T. (2022). Analysis of herding behavior in individual investor portfolios using machine learning algorithms. *Research in International Business and Finance*, 62: 101740. <https://doi.org/10.1016/j.ribaf.2022.101740>
- [20] Jaradat, A., Alhussian, H., Patel, A., Fati, S.M. (2020). Multiple users replica selection in data grids for fair user satisfaction: A hybrid approach. *Computer Standards & Interfaces*, 71: 103432. <https://doi.org/10.1016/j.csi.2020.103432>
- [21] Petit, M., Ray, C., Claramunt, C. (2010). Algorithme de recommandation adaptable pour la personnalisation d'un système mobile. In 6èmes Journées Francophones Ubiquité et Mobilité: UBIMOB 2010, pp. 59-62.
- [22] Han, S.H., Mutahira, H., Jang, H.S. (2023). Prediction of sensor data in a greenhouse for cultivation of paprika plants using a stacking ensemble for smart farms. *Applied Sciences*, 13(18): 10464. <https://doi.org/10.3390/app131810464>
- [23] Anassin, M.C., Aka, B., Brou, M.K. (2018). Data extraction in the warehouse: A quality-based approach. *European Journal of Scientific Research*, 150(1): 14-20.
- [24] Das, S., Narula, P., Sarkar, K. (2020). Design of intermittent rainfall-pattern for structures with gridded data: Validation and implementation. *Journal of Building Engineering*, 27: 100939. <https://doi.org/10.1016/j.jobe.2019.100939>
- [25] Charalambidis, A., Troumpoukis, A., Konstantopoulos, S. (2015). SemaGrow: Optimizing federated SPARQL queries. In *Proceedings of the 11th International Conference on Semantic Systems*, Vienna Austria, pp. 121-128. <https://doi.org/10.1145/2814864.2814886>
- [26] Cheng, Y., Zhou, K., Wang, J., Cui, S., Yan, J., De Maeyer, P., Van de Voorde, T. (2022). Regional metal pollution risk assessment based on a big data framework: A case study of the eastern Tianshan mining area, China. *Ecological Indicators*, 145: 109585. <https://doi.org/10.1016/j.ecolind.2022.109585>
- [27] Singh, V., Chen, S.S., Singhania, M., Nanavati, B., Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries-A review and research agenda. *International Journal of Information Management Data Insights*, 2(2): 100094. <https://doi.org/10.1016/j.ijime.2022.100094>
- [28] Morón-López, J., Rodríguez-Sánchez, M.C., Carreño, F., Vaquero, J., Pompa-Pernía, Á.G., Mateos-Fernández, M., Aguilar, J.A.P. (2020). Implementation of smart buoys and satellite-based systems for the remote monitoring of harmful algae bloom in inland waters. *IEEE Sensors Journal*, 21(5): 6990-6997. <https://doi.org/10.1109/JSEN.2020.3040139>
- [29] Mohamed, C., Rizzi, S., Rachid, C. (2017). Enabling self-service BI on document stores. In *Proceedings of the EDBT/ICDT 2017 Joint Conference*, Venice, Italy, pp. 1-10.
- [30] Soares, C.G., Santos, T.A. (2023). Advances in maritime technology and engineering. In *Proceedings of the 7th International Conference on Maritime, Technology and Engineering (Martech 2024)*, Lisbon, Portugal, pp. 714.