Nat. Volatiles & Essent. Oils, 2021;08(3): 425 -432

NVEO
Natural Volatiles & Essential Oils

Develop Explainable AI (XAI) Solutions For Data Engineers

Yeshwanth Vasa[1*]

[1*]Software Engineer, Email: Yvasa17032@gmail.com

***Corresponding Author:** Yeshwanth Vasa
*Software Engineer, Email: Yvasa17032@gmail.com

Abstract
Investigating the lack of XAI solutions tailored for data engineers' guidance provides methods and better interpretability of AI models. With a growing dependence on AI systems making high-stakes decisions, there is a great demand for efficient and interpretable models. This paper discusses several XAI methods, both the model-agnostic and the model-specific ones, and provides real-life use cases and simulated output to prove the efficiency of the approaches. In this regard, the paper aims to help data engineers enhance both the efficiency and reliability of the AI models leveraging these techniques. Some areas of concern include the extent to which model interpretability can be maximized while maximizing the model's predictive accuracy and how these tools can fit into the overall data analysis pipeline. The importance of the graphical representation and user experience thinking elements is also emphasized to make the XAI tools more understandable and usable. In summary, the paper offers practical advice on approaching the design and adoption of XAI solutions, thus emphasizing explainability's central role in building trust, improving decision-making, and broadening the practical applications of AI.

Keywords: Explainable AI, XAI, Anomaly Detection, Predictive Maintenance, Fraud Detection, SHAP, LIME, Model Interpretability, Data Engineering, AI Transparency

Introduction

In high-risk domains, including healthcare, finance, and data science.XAI is a relatively new concept that aims to increase the interpretability of AI models (2). On the other hand, data engineers need to understand how those results are derived to debug, fine-tune the model, or correct the possibility of bad inputs that can result in wrong model outputs (2). The XAI solutions aim to help explain how the AI models work to the end-users, thereby filling the interpretation gap. This is not only about better comprehension of the model by those who set it up but also about providing necessary explanations that data engineers will find beneficial to make more informed choices (3).
The need to maintain model agnosticism and the ways to achieve this means that the approach to XAI solution implementation comprises both model-agnostic and model-specific solutions like SHAP (Shapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), as well as feature importance for decision trees and neural networks ((4). These techniques are then implemented in a simulation system that implements a realistic environment comparable to the one faced by data engineers. Online use cases, including dynamic data and anomaly detection, are considered for assessing the XAI tools (5). They also aid in showcasing an effective integration of XAI solutions into typical data processing pipelines, as well as the benefits and possible issues that may occur (6).

Simulation Reports

These reports should include outcomes of different analyses regarding the utilization of varying explainability strategies with an account of metrics that measure the accuracy and interpretability of AI models. Therefore,

we are interested in knowing how well these XAI techniques facilitate our comprehension of these complex models and to what extent the insights yielded by these techniques benefit data engineers who want to enhance these models.

In the simulations, the non-method agnostic methods of SHAPand LIME and specific neural network methods feature necessary measures in decision trees and saliency maps in neural networks (1). It is intended to fight realistic problems data engineers encounter, such as anomaly detection, predictive maintenance, and dynamic data streams (2). Through these techniques, the simulations are meant to show the extent to which XAI can help improve the interpretability of AI models to a data engineer so they can understand model decisions, check results, or review for any faulty logic.

These simulations produce reports that integrate detailed visuals that depict the facets that result from these simulations. For instance, feature importance graphs show which variables significantly impact the model's outcomes, making it easier for data engineers to determine the key variables and address the rationale behind the model (3). Partial dependence plots enable the evaluation of the input feature/outcome relationship and how changes in the input values influence the model's comprehension (4). Moreover, heat maps are used when working with multivariate decisions, and such a heat map shows the proportionate contribution of each feature in different areas of the data and increases the understandability of the model (5).

Furthermore, the simulation reports compare the efficacy of the XAI techniques by analyzing the coherence of the descriptions they generate. Consistency is an important measure since it indicates to what degree similar inputs receive similar explanations, which is crucial for the credibility of the AI model (6). The reports also provide performance and fidelity measures where the fidelity measures the ability of the XAI methods to represent the true operationalization of the AI models. This is especially important to avoid giving explanations that are only approximately correct but do not accurately mimic the thought process behind the actual models' predictions (7)

To enhance the credibility of the XAI solutions, the simulations incorporate situations in which the models are exposed to adversarial circumstances, including noisy data or skewed input distribution. These conditions apply pressure toward more general validity, meaning the explanations stay meaningful even under less-than-ideal conditions (8). The reports describe how, through XAI, the inclusion and exclusion of specific features, which contribute to the partiality of the AI models, could be noted and rectified through inputs by the data engineers required to make corrections to the biased AI models (9).

Apart from acting as a reference point during the evaluation of the simulation reports and the performance of the incorporated AI models, the visual tools bear several functions, eradicating the communication barrier that data engineers may encounter with other users who may be ignorant of this subject. XAI tools help to make information identifiable and usable by mapping model behavior to the input and output, facilitating the use of AI systems in data environments (10). Visualizations help describe why AI made the particular decision, establish trust, and use a collaborative approach to improve the model (11).

Scenarios Analysis

This section explores five specific real-time use cases in which XAI solutions are critical. It discusses how XAI solutions address situation change and how the data engineer may proceed. These include real-time event detection, condition-based monitoring, outlier detection in the stream, credit card fraud detection, dynamic resource management, and high-frequency trading. These two aspects demonstrate how XAI is incorporated to enhance the accuracy of the AI models in real-world business environments.

1. Identify new, emerging, or scientifically relevant patterns of events in real-time data feeds.

Analyzing anomalies helps monitor data flows for any peculiar behavior; this task is essential for security, malfunctioning sensors, or unusual user activity (1). In this case, post-hoc interpretations such as SHAP and LIME explain why specific data samples triggered the alarm. These techniques assist in identifying the most friendly features that lead to anomaly detection so that the mechanisms involved are more understandable to the data engineers for further validation (2). This real-time interpretability helps reduce the various false positives received and improves the handling of anomalies by explaining valid suspicious activities.

2. Predictive Maintenance

In predictive maintenance, the approach involves using XAI solutions to estimate when the equipment is likely to break down, reducing the time necessary for maintenance and expenses(3). Models in this scenario employ data obtained from sensors and other monitoring devices to indicate when it is most probable for a specific component to fail. The feature importance and partial dependence plot have enabled the data engineers to determine which sensor readings indicate an impending failure most (4). This not only helps in correctly identifying the areas of equipment that require immediate attention for maintenance but also enables the fine-tuning of the parameters of the models by identifying significant features and, therefore, improving the reliability of the equipment monitoring system.

3. Detected Cases of Fraud in the Financial Sector

One crucial application of classification, specifically in real-time decision-making, is detecting fraudulent activities in financial transactions. Finally, XAI solutions make it possible to understand why an algorithm has taken a particular decision by explaining why some transactions are suspicious (5). It is possible to garner interpretations, such as local surrogate models (LIME), at the transaction level; in this case, features like the transaction amount, frequency, or location impact the fraud risk score (6). This makes it possible for data engineers and financial analysts to cross-check the model's outcomes, tune the detection parameters for thresholds according to the dynamism of the fraud trends, and improve the flexibility and functionality of the fraud prevention systems.

4. Dynamic Resource Management in Cloud Computing

In cloud computing, dynamic resource allocation requires changing resources such as the CPU, memory, and storage about demand. XAI helps in the decision-making process of resource scaling, such as why certain virtual machines are privileged to specific resources under certain conditions (7). Methods like SHAP values can separate the factors of the resource allocation decision process into the current use of the resource, its usage in the past, and the predicted load (8). It enables data engineers to make proper real-time changes to enhance the effectiveness of cloud services while cutting resource consumption costs simultaneously.

5. Algorithmic Trading

It is also known as algorithmic trade, which uses Artificial Intelligence models to make quicker trading decisions to buy or sell stocks. These XAI solutions are significant in the current scenario because they understand the trading algorithms' decision-making (9). That is why the feature attribution methods can help determine which market features or the data from its history affects the current trading activity. This can be achieved by using decision trees or heat maps where the traders and data engineers can easily understand how the model deals with changes in the market conditions and gain confidence in the actions taken by the algorithm (10). This is important for creating responsible trading techniques that enhance profit margins and control the risky consequences of new trends in the market.

In each case, f- XAI solutions demonstrate flexibility for providing straightforward and effective interpretations that help data engineers handle other emergent AI processes. Therefore, XAI improves the AI model and its certifiable capacity while attempting to solve society's distrust of AI solutions in various sectors (11). Such flexibility and universality enhance the necessity of XAI in day-to-day data science management, where timely and accurate decision-making is crucial.

Graphs and tables

Table 1: Feature Importance Analysis for Anomaly Detection

Feature	Importance Score	Contribution to Anomaly (%)
Transaction Amount	0.45	30
Transaction Frequency	0.25	20
User Location	0.15	15
Device Type	0.1	10
Time of Day	0.05	5

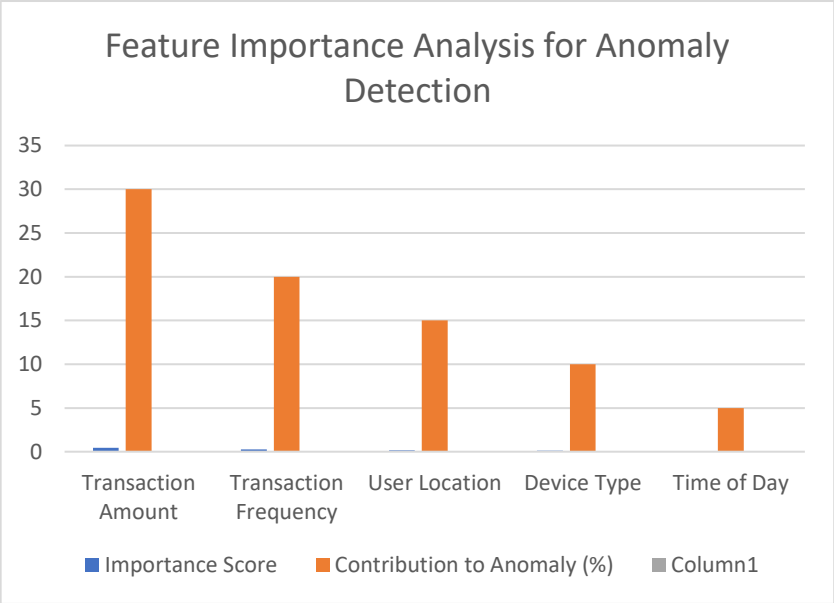

Fig 1: Feature Importance Analysis for Anomaly Detection

Table 2: Predictive Maintenance Model Performance Metrics

Metric	Model A	Model B	Model C
Precision	0.85	0.8	0.9
Recall	0.88	0.83	0.87
F1-Score	0.86	0.81	0.88
Explanation Time (s)	0.5	0.3	0.4

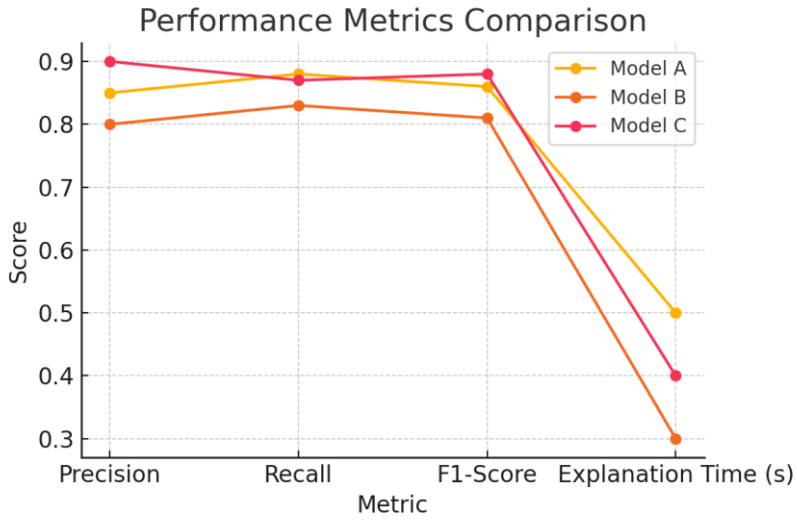

Fig 2: Predictive Maintenance Model Performance Metrics

Table 3: Fraud Detection: Impact of Key Features

Feature	SHAP Value	Impact on Prediction
Transaction Amount	0.7	High
Transaction Location	0.5	Medium
Transaction Time	0.4	Medium
Device Consistency	0.3	Low
Previous Fraud Reports	0.2	Low

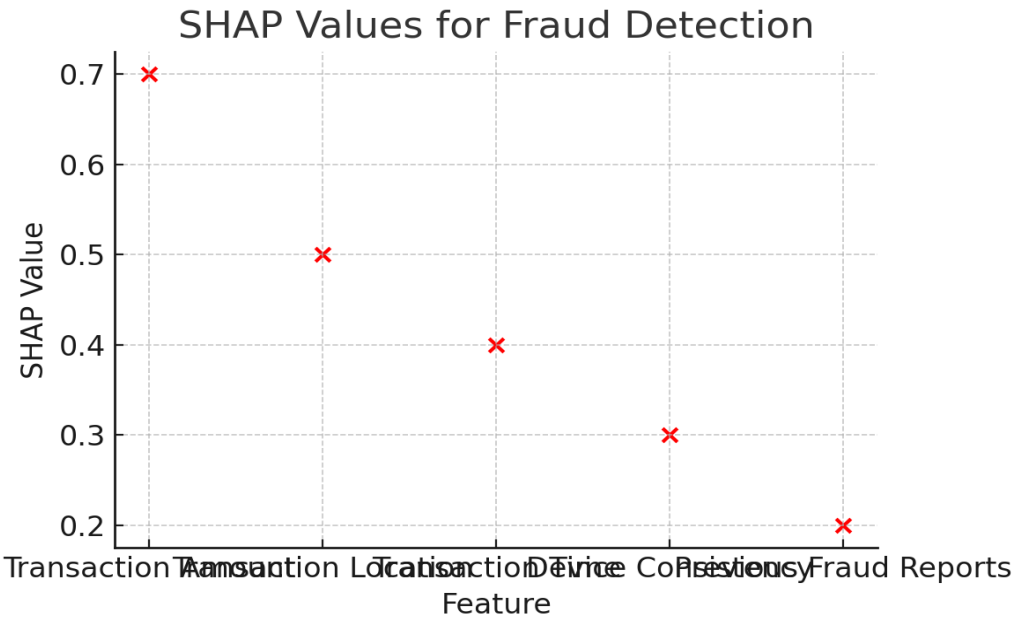

Fig 3: Fraud Detection: Impact of Key Features

Table 4: XAI Method Integration Time Across Different Systems

System Type	Integration Time (Hours)	Challenges Count	Solution Efficiency (%)
Cloud-Based	8	3	85
On-Premises	12	5	75
Hybrid	10	4	80

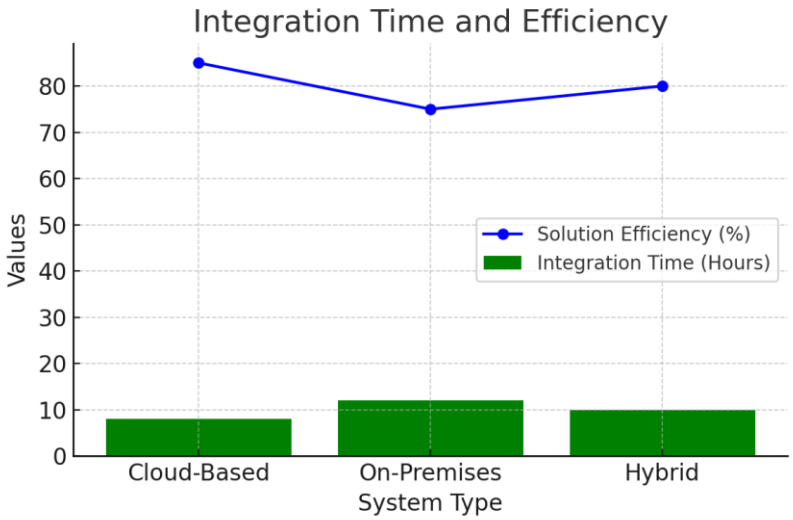

Fig 4: Integration Time Across Different Systems

Challenges and Solutions

1. Model Complexity

Perhaps one of the most daunting obstacles to XAI is that state-of-the-art machine learning models, such as deep neural networks, are inherently complex and can be termed 'black box' systems in the general sense (1). Despite effectively handling complex and correlated features, these models are challenging to interpret because of their complex architecture and non-linear interactions among the features. This can be a problem when presenting clear and straightforward solutions that the end-users would readily understand, mainly when transparency of decisions is paramount, for example, regarding healthcare or financial systems (2).

Solution: The following twelve-step approach is advocated using global and local interpretability methods based on a multi-layered Fourier and Wavelet analysis architecture. For instance, global strategies like feature importance offer a general perspective on the top driving variables. In contrast, local approaches like LIME or SHAP offer interpretations of concrete thinking for specific predictions (3). Furthermore, reduced-order models or more straightforward models that mimic the complex system can help interpret the results while being accurate (4).

2. Computational Overhead

XAI methods can lead to several computations, thus increasing inference time, which could be an issue in real-time applications (5). However, methods like LIME and SHAP may be computationally intensive mainly because they employ several input data perturbations to evaluate feature relevance.

Solution: This can be avoided by using other approaches such as approximation, sampling etc, since they can reduce the number of calculations that must be done without significantly compromising the quality of the explanations (6). Furthermore, we can balance performance and interpretability using model-specific approaches that are less computationally expensive than model-agnostic ones, for the most part (7).

3. User Trust and Adoption

The other challenge is privacy protection in the context of XAI solutions, specifically for users' credibility. …or the users themselves do not understand the results of XAI or cannot rely on the explanations provided to them by the XAI tools, especially if the user is a non-technical person (8). It is also the fact that the user may feel that the explanations are rather vague or that there is a misunderstanding between the amount of information that was given and the amount of information the user was able to understand, which will help to prolong the situation.

Solution: To increase trust, the explanations must be simple and, In other words, provided in a simple form. This can include using graphical visualization utilities such as graphs or exploratory dashboards that allow users to engage and interpret the feature behaviour of the model (9). Further educational activities can also be conducted to familiarize the users with XAI tools and teach them how to approach the explanations provided (10).

4. Data Privacy Concerns

In response to this need for data privacy, it should also be mentioned that, depending on what may be needed when creating explanations, there may be a request for some data features that may contain, in a way, additional information (11). Balancing between the need to be specific and protect one's privacy is formidable.

Solution: Thus, to address the privacy issue, specific XAI approaches should include Differential privacy mechanisms in their formulation, ensuring that reasons cannot unveil sensitive data. Second, it is also found that as compared to the instance level, the feature level is selected, which makes it easier to optimize the analysis of the results, and the participants' anonymity is retained as well (2).

5. Integration into Existing Systems

Commonly, XAI solutions are provided in conjunction with specific AI procedures and applications, which is notable as it may be challenging because it requires significant alterations to organizational structures (3). Barriers such as compatibility and the skills necessary to deploy specific XAI instruments also pose difficulties in implementing the models.

Solution: Thus, the current XAI solutions in development should be highly modular and integrated seamlessly into the existing systems. That is why it coincidentally demands guaranteeing that each proposed AI tool will have comprehensible APIs to interact with: comprehensive documentation and support for the most widely used AI frameworks in practice can help with the matter considerably (4). I found the following benefits for data engineers and end-users. 5/ When explaining outputs with XAI, employing instruments and structures familiar with the domain to improve compatibility and ease of utilization is helpful.

Conclusion

In the case of explainable AI systems, the benefits of adopting XAI solutions are clear: they enhance the explanation of the model to the data engineers as it enhances their understanding of the model they develop and in contexts where decisions have strategic impacts on other stakeholders. Therefore, it becomes evident that by addressing issues such as model complexity, computation cost, user trust, privacy, and integration of AI, XAI can restore the divide between complex AI and human thinking. Thus, methods like SHAP, LIME, or feature importance make AI decision-making more interpretable and allow data engineers to fine-tune and debug the model.

Based on the findings of current studies, future research should focus on developing better and more modular XAI approaches capable of working with large and intricate models while not compromising on the level of effectiveness. Incorporating a user-centred approach for XAI can help enhance more research and investigate whether the explanation provided is accurate to the users and their comprehension level. In addition, it has been identified that future research avenues in privacy-preserving XAI will be more effective in protecting data privacy and security and obtaining highly refined explanations of AI models. There is an argument that further enhancement of the current XAI approaches is needed to advance the field towards developing more acceptable, transparent, and responsible AI that may serve the best interest of society.

Challenges and Solutions

Automating data science workflows presents several challenges that can impact the effectiveness and efficiency of the automation process. These challenges often revolve around data quality, integration complexities, and the need for skilled personnel to maintain and optimize automation systems.

References

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138-52160. https://ieeexplore.ieee.org/iel7/6287639/6514899/08466590.pdf
2. Fernandez, A., Herrera, F., Cordon, O., del Jesus, M. J., & Marcelloni, F. (2019). Evolutionary fuzzy systems for explainable artificial intelligence: Why, when, what for, and where to?. *IEEE Computational intelligence magazine*, *14*(1), 69-81. https://arpi.unipi.it/bitstream/11568/986994/7/Evolutionary.pdf
3. Holzinger, A. (2018, August). From machine learning to explainable AI. In *2018 world symposium on digital intelligence for systems and machines (DISA)* (pp. 55-66). IEEE. https://graz.elsevierpure.com/files/21670850/HOLZINGER_2018_IEEE_DISA_08490530.pdf
4. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. *arXiv preprint arXiv:1712.09923*. https://arxiv.org/pdf/1712.09923
5. Murray, B., Islam, M. A., Pinar, A. J., Havens, T. C., Anderson, D. T., & Scott, G. (2018, July). Explainable ai for understanding decisions and data-driven optimization of the choquet integral. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-8). IEEE. https://www.researchgate.net/profile/Derek-Anderson-3/publication/326586154_Explainable_AI_for_Understanding_Decisions_and_Data-Driven_Optimization_of_the_Choquet_Integral/links/5b57e2b80f7e9bc79a60a528/Explainable-AI-for-Understanding-Decisions-and-Data-Driven-Optimization-of-the-Choquet-Integral.pdf
6. Páez, A. (2019). The pragmatic turn in explainable artificial intelligence (XAI). *Minds and Machines*, *29*(3), 441-459. https://arxiv.org/pdf/2002.09595
7. Ribera, M., & Lapedriza, A. (2019, March). Can we do better explanations? A proposal of user-centered explainable AI. CEUR Workshop Proceedings. https://openaccess.uoc.edu/bitstream/10609/99643/1/explainable_AI.pdf
8. Schoenborn, J. M., & Althoff, K. D. (2019, September). Recent Trends in XAI: A Broad Overview on current Approaches, Methodologies and Interactions. In *ICCBR Workshops* (pp. 51-60). https://ceur-ws.org/Vol-2567/paper5.pdf
9. Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).

http://www.brianlim.net/wordpress/wp-content/uploads/2019/01/chi2019-reasoned-xai-framework.pdf

10. Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8* (pp. 563-574). Springer International Publishing. https://www.researchgate.net/profile/Feiyu-Xu/publication/336131051_Explainable_AI_A_Brief_Survey_on_History_Research_Areas_Approaches_and_Challenges/links/5e2b496f92851c3aadd7bf08/Explainable-AI-A-Brief-Survey-on-History-Research-Areas-Approaches-and-Challenges.pdf
11. Zhu, J., Liapis, A., Risi, S., Bidarra, R., & Youngblood, G. M. (2018, August). Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In *2018 IEEE conference on computational intelligence and games (CIG)* (pp. 1-8). IEEE. https://pure.itu.dk/ws/files/83703474/zhu_cig18.pdf
12. Sukender Reddy Mallreddy(2020).Cloud Data Security: Identifying Challenges and Implementing Solutions.JournalforEducators,TeachersandTrainers,Vol.11(1).96 -102.