Federated Analytics: Insights Without Centralized Data

Dheeraj Vaddepally

dheeraj.vaddepally@gmail.com

Abstract:
Federated analytics offers a paradigm shift in analysis. This is due to the availability of insights of distributional data minus the burden of centralized data storage. Generally, this paper presents principles and methodological approach as used in federated analytics of statistical analysis of patterns detection within federated statistical analytics, an application that covers various strict demands of data secrecy. Federated analytics directly tackles critical concerns that include privacy and risks in the centralized management of data through federated learning, secure multi-party computation, and differential privacy. This work dives into practical applications and emerging opportunities within these domains, showing how federated analytics can provide powerful, privacy-preserving insights in compliance with the legal and ethical standards. The paper talks about federated analytics' present-day limits and future as of now focusing both on scalability and security aspects where the scope is needed with further collaboration towards decentralized data analysis.

Keywords: Federated analytics, statistical analysis, pattern detection, federated learning, privacy, data secrecy, secure multi-party computation, differential privacy, decentralized data analysis, scalability.

I. INTRODUCTION

Due to the surging demand for insights from data, industries heavily relied on central data storage and analysis. Yet, this model raises critical questions about data privacy, security, and compliance-questions that are generally sensitive in such domains as health care and finance. Central data from many sources increase the danger of breaches and unauthorized access and thus create problems regarding individual privacy and organizational confidentiality.[1] Federated analytics is now a promising means of answering such obstacles in enabling the analysis of data without requiring it to be centrally located.

Federated analytics applies decentralized sources of data so that insights could be generated with collaborative computation from multiple nodes that keep raw data local to the participating entities. [1] This minimizes the opportunity for leakage of confidential information and will help to follow strict data privacy regulations. It is based on the distributed computing framework and advanced statistical analysis techniques forming the basis for federated analytics, where patterns can be detected, predictive models built, and decisions made without the compromise of integrity or privacy of the data.

The paper outlines the principles that are to federated analytics, including what it can bring to statistical analysis and pattern discovery in sensitive domains that include healthcare and finance, for example. This paper goes on to discuss some of the enablers, including federated learning, secure multi-party computation, and differential privacy, all of which ensure that such analytics may be performed while preserving privacy. Shed light is thrown on federated analytics so that it explains how it would change the character of traditional paradigms to data analysis to preserve privacy with high-quality insights, which otherwise will not be yielded. [2]

II. BACKGROUND

Usually, centralized data analytics are based on the collection of large amounts of information from sources that are spread across various other locations and then processed and analyzed in a single, centralized location. This practice has been one of the underpinnings of data-based decision-making. It has empowered

organizations to undertake extensive analyses and yield valuable insights into their operations and data.[2] However, centralized analytics has big issues, especially on data privacy and security. Data that is collected and stored at one central point becomes an attractive focal point for cyber-attacks. Such attacks can compromise sensitive information. The approach may also contravene data privacy legislation, such as the GDPR, which limits the way data is collected, stored, and disseminated. Further, centralized data systems are costly and inefficient as large resources in terms of storage, maintenance, and computer power are required. [1]

Federated analytics offers a decentralized approach that mitigates many of these challenges by allowing data analysis to occur at the local level without having raw data transfer to a central repository. Federated analytics allows the analysis of data by different parties in coordination with each other while keeping control of their local datasets instead of pooling data in a central server. [3] This method enables insights to be derived from distributed sources while preserving privacy and reducing the risks associated with centralized data collection. It allows models to be trained on local data, only model updates or aggregated results shared across the network and thereby ensures that sensitive data never leaves its source, thus permitting real-time analysis and insights from diverse, distributed datasets.

Federated learning is a subfield of machine learning; federated analytics is enabled by this aspect. In federated learning, the models of machine learning are developed across decentralized sources of data. The model gets trained locally within the individual devices and the servers, whereas aggregation and updating the model parameters happen in a central server.[1] It won't allow one single point to be transferred or shared in any way; this avoids highly accurate models without having it as a sacrifice on privacy or need to centralize the sensitive data. Federated learning securely collaborates entities in training various models with totally different datasets to their credit hence being convenient for industries engaged in sensitive as well as personalized data.

Most key technologies support successful federated analytics implementations: the core element of federated analytics is an edge computing methodology, meaning in other words-performing data processing and computation on the spot with the generation, close to that source, probably on a personal device.[2] It reduces latency and requires fewer needs to transfer massive chunks of data, thus realizing real-time analytics in decentralized scenarios. Another major technology is SMPC, with which multiple parties can jointly compute results without breaching confidentiality in their individual datasets. However, the differential privacy techniques bring yet another technique, such as adding noise to the aggregated results so that the analysis cannot leak what information it has about any individual data point. In other words, SMPC protocols are protocols which enable the processing of data in such a way that no participant may access the raw data of another; hence privacy is also preserved in the analysis process. Together, these technologies do make for quite robust and secure federated analytics performing environment that offers a very powerful alternative to traditional data analysis methods through centralization.

III. FEDERATED ANALYTICS METHODS

Statistical methods such as regression, classification, and clustering are applied locally on distributed datasets to discover meaningful patterns and insights in a federated analytics environment. For instance, regression will train the model across multiple devices or servers that each handle its local data set and aggregate the model parameters to generate a global prediction. On classification, federated analytics enables the training of classifiers on the local datasets. The updates obtained from these are then disseminated among various nodes, which enables fraud detection or sentiment analysis without exposing the raw data outside its original source.[4] Clustering techniques can also be applied in federated environments to group similar data points locally. Aggregated results then reveal broader patterns across the distributed datasets, such as customer segmentation or anomaly detection in sensor data. These statistical methods, when used federatively, make possible the efficient detection of trends and correlations without centralizing data.

Fig. 1. FA requirements

The most common techniques in discovering patterns when federated analytics is involved usually make use of integrated models of statistics and those with machine learning algorithms that learn from dispersed sources of data. In particular, the most popularly used are decision trees, support vector machines, and neural networks. For instance, decision trees can be constructed across multiple data sources in which local models are learned and then aggregated into a global decision-making tree. [4] Federated support vector machines can be learned in a distributed fashion for the classification of data points without violating privacy. Neural networks, including deep learning models, are applied more and more in federated environments for image recognition and natural language processing tasks. In these cases, federated learning would enable this data to stay local at their source while the parameters of the models are shared and allow the network to learn in a global context from diverse sets of inputs to the data. The big deal with this particular technique in this federated domain is that these models get more accurate over time with origin data without selling out sensitive data.

Federated analytics may provide a myriad of benefits; however, this approach is facing numerous challenges and limitations. Majorly, there is model convergence because federated systems train on heterogeneous datasets which might have differently distributed data, making it quite challenging to find an agreement with the models being trained locally. This can lead to slow convergence and suboptimal performance. Another challenge includes communication overhead; the decentralized nature of federated analytics requires quite frequent exchanges of model updates, which can be resource-intensive. This becomes especially problematic when dealing with large-scale datasets or a large number of nodes. Another challenge is handling heterogeneous data sources: data from different nodes may differ in format, quality, and distribution, making it challenging to apply uniform models across all sources. To mitigate these issues, new techniques for data normalization, federated optimization algorithms, and strategies for reducing the communication cost are being developed to enhance the efficiency and accuracy of federated analytics.

IV. APPLICATION OF FEDERATED ANALYTICS

Federated analytics holds very great promises in all sectors, especially in domains whose major concern is the safety and privacy of personal data.

- Federated analytics revolutionizes the way one analyzes the medical data of patients. The hospitals, research institutions, and health service providers get an opportunity to be at work together in the studies and clinical trials without losing the patient's data confidentiality. Medical institutions can train machine learning models on local patient records and medical imaging data. [5] The insights that are discovered by these models can then be shared and aggregated to generate a more comprehensive understanding of diseases, the effectiveness of treatment, and public health trends. This makes possible the use of sophisticated predictive models without violating such strong data protection laws as HIPAA, so personal health information is never shared between organizations.

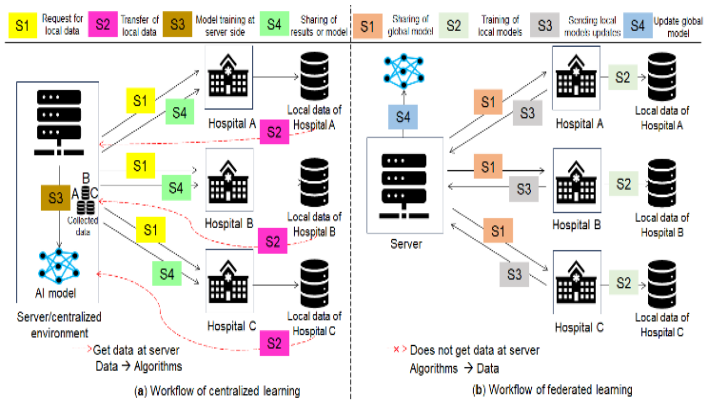

Fig. 2. Analysis of Federated Learning Paradigm in Medical Domain

- In banking and finance, federated analytics applies over areas such as fraud detection, credit scoring, or market prediction - all areas that involve privacy criticality over their customers. Such insights and model parameters can be shared while keeping all sensitive financial information secure. The advantages of federated analytics in this domain are immense; one major one is that local analyses based on transactions with clients can be conducted by banking and other institutions partnering on the creation of machine learning-based fraud-detecting models for credit risk. It would facilitate market analysis by the financial institutions for corresponding investment strategies based on decentralized data such as market trends, trading volumes, and customer behavior patterns.

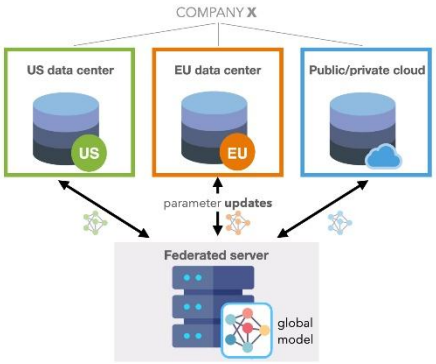

Fig. 3. Analysis of Federated Learning Paradigm in Finance

- Sharing insights without revealing private or classified data in sensitive domains like government agencies, defense, and insurance becomes possible through federated analytics. This will enable governments to carry out analysis on data pertaining to national security, like identifying cyber threats or forecasting the occurrence of a natural disaster, without exposing sensitive citizen data. Federated analytics within the insurance industry could contribute to improving the development of appropriate risk models and claims prediction algorithms by insurance companies with the data of customers remaining confidential. Other domains that federated analytics can contribute are agriculture and energy, where distributed sensor sources like weather stations or smart meters are analyzed to optimize operation and resource usage without sharing sensitive or proprietary information.

In a nutshell, federated analytics provides organizations engaged in privacy-sensitive businesses with an ability to cooperate in tapping the powers of distributed data while opening precious insights unlocked safely and securely. This will keep extending to even more sectors once models of federated learning develop, along with corresponding technologies, such as is going to be done with federated analytics in relation to decentralizing data analytics work.

V. PRIVACY AND SECURITY CONSIDERATIONS

Federated analytics greatly benefits within the scope of privacy, due to supporting analytics without a centralization need, especially for sensitive data. However, the challenge has been the aspect of concerns

concerning privacy within heavily regulated sectors and industries like finance and healthcare. One of the biggest privacy problems when it comes to using federated analytics revolves around maintaining that data will still be kept private even though its use is for training in developing machine learning. Unendingly, in contrast to the case of centralized analytics where raw data are moved to a server central for processing, federated analytics ensures that data stays local to the source. However, during training, model parameters updates get shared throughout the network. This raises a concern that even though access to raw data is not direct, an adversary could reverse-engineer these updates for inferring some sensitive information. This risk is mitigated using federated learning, which avoids the revelation of individual data points using techniques such as local differential privacy and secure multi-party computation.

Model updates are shared in federated learning in aggregated forms instead of raw data, which specifically helps mitigate privacy issues. This means that federated learning prevents the direct access to sensitive information by focusing on the transmission of model parameters rather than individual data points and yet allows the model to learn from a broad set of data sources. Federated learning can be augmented with local differential privacy, which adds noise to the model updates in such a way that any single update does not reveal private data. This would ensure the privacy of those whose data was used to train the models, even when a model is trained across multiple institutions or devices.[5]

Various security mechanisms have been implemented in federated analytics so that the data being transferred would not be misused. The most common among them is encryption. It makes sure that data is kept confidential while transferred through the network. Such a model updates and guards the aggregation results using encryption protocols such that the interception of communication will never make that information readable without its decryption key. The other federated analytics mechanism key is called secure aggregation. This mechanism is making sure that the aggregation of data from different participants is done in such a way that no participant's data can be accessed. Instead of broadcasting raw model updates, participants encrypt and send the updates to a central server, where the server aggregates the updates securely, thus updating the global model without ever seeing the individual updates. Differential privacy is also a critical component in securing federated analytics since adding noise or randomness to the sum of the model updates aggregated makes the approach stronger against attackers. This ensures that even if some malicious actor is trying to reverse-engineer the updates, extraction of individual data points is almost impossible.

While federated analytics offers benefits of great magnitude regarding privacy and security, such technologies also raise their associated legal and ethical dilemmas. From a legal perspective, organizations need to ensure that federated models for analytics meet all the data protection regulations. The U.S. has the Health Insurance Portability and Accountability Act, while the EU has the General Data Protection Regulation. The rules of those Acts do indeed set strict standards concerning data privacy. Privacy entails having informed consent by individuals whose data is used, as well as a right to be forgotten. Design almost automatically mitigates part of the problem because it is not going to centralize people's data, but then this would still result in legal troubles on cross-border data sharing or trying to see whether the aggregated data comply with such laws.[6]

Federated analytics also has to solve ethical questions, too, in balancing data sharing with preserving privacy. While the technology itself may indeed enhance privacy, a good implementation will be sensitive to the potential biases in the data, model fairness, and the risk of unintended consequences. For example, federated models must be continuously monitored to avoid reinforcing pre-existing biases or acting discriminately in sensitive application domains such as health and finance. Other than the technical requirements of federated learning, the process for federated learning should be transparent and all stakeholders understand what is being learned through the models and types of data involved. [6] Ethical considerations exist as far as predictive policing or hiring algorithms, for example, as the effects of biases or inaccuracies within models can severely affect society.

VI. POTENTIAL AND FUTURE DIRECTIONS

A. Scalability

Federated analytics could scale well and process massive amounts of distributed data spread over thousands of devices or institutions. The approach to federated analytics contrasts with other methods of analytics in that the former does not centralize the data; thus, data remain distributed, with lower communication costs and decreased exposure to the risk associated with mass-scale data gathering.[7] Effective scaling of federated analytics depends on how one overcomes challenges that pertain to communication overhead, heterogeneity of devices, and quality differences among data sources.

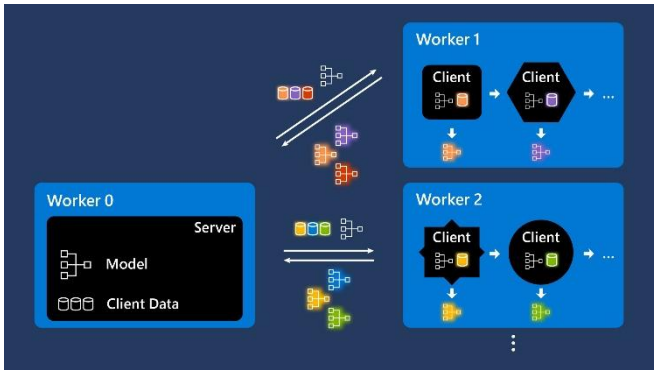

Fig. 4. Scalable Federated analysis simulation

The primary concern in terms of scalability would be model synchronization. As the number of devices or nodes involved in learning grows, model updates can get challenging to be synchronized. This problem is currently being addressed through solutions such as asynchronous updates and batch updates, which help mitigate communication overhead. The edge computing plays a vital role in scaling federated analytics as it will be possible to process data at the device level, hence less dependency on central servers and quick processing of analytics. For applications across healthcare, finance, and telecommunication, it is crucial that federated models can scale without affecting the speed or accuracy in order to process real-time data for millions of people.[7]

B. Future Research Directions

Federated analytics is still a relatively new field with several very important areas that require research and development.

- Efficiency Scalability: Improving communication efficiency and cutting through latency with scalings will be an essential factor in federated analytics. Techniques such as model compression, which makes the models smaller in size without losing the accuracy of the same, is promising efficiency-enhancing avenues. Sparse models and more efficient aggregation protocols decrease the amount of data exchanged between devices and servers.
- Model Accuracy: federated analytics only face one large challenge-that is the heterogeneity of the source of data. Personalized federated models with a knowledge that will allow for particular data distributions among different nodes so as to give better accuracy along with improving the robustness to adversarial attacks of federated models by robustly accurate against such malicious parties or faulty sources of data.
- Improvements in Security: The federated analytics framework gives utmost importance to security at places because it deals with sensitive domains. Increased research in the area of secure aggregation protocols and federated learning with homomorphic encryption will improve the privacy of model updates. Techniques for differential privacy that strengthen security without reducing the utility of data also help in reducing privacy concerns. Handling adversarial models and data poisoning is another area of the federated learning architecture focused on system integrity enhancement.
- Interoperability: Because federated analytics cuts across industries and data formats, interoperability between different systems and platforms is also a critical issue. Future research should focus on the standardization of protocols and interoperability across various federated systems to allow collaboration across organizations with minimal effort.

C. Adoption and Implementation

All the industries that deal with sensitive privacy fields, especially healthcare and finance, will be taking on federated analytics.

1. Adoption Challenges

- Implementation: It requires tremendous infrastructure: with devices for edge computing, safe data aggregation, and communication network; hence the task is somewhat of a nightmare to the not so well-prepared organizations lacking adequate resources and expertise.
- Data Quality and Heterogeneity: Since federated analytics are distributed in nature, quality and form could differ significantly based on the source of data. Discrepancies in data might impact models and fail to carry out their analytics effectively, so data compatibility would be a severe limitation to more popular adoption of this system.
- Legal and Regulatory Challenges: Though federated analytics helps to advance privacy, regulatory aspects of data sharing across different institutions are challenging. For example, laws under the GDPR, HIPAA, etc., are varied, and organizations are accountable for full compliance, thus making adoption more hard, especially in multi-jurisdictional or border crossing scenarios.

2. Opportunities for Adoption:

- Healthcare: Federated analytics has tremendous potential to transform healthcare through collaborative research and model development without violating patient privacy. Federated analytics can accelerate medical discoveries and improve patient outcomes without centralizing sensitive medical data by sharing insights across hospitals and research institutions. This could be especially impactful in medical imaging, genomic data analysis, and drug discovery.
- Finance: In the financial industry, federated analytics can be applied where banks and financial institutions can detect fraud, assess risk, and predict market trends without disclosing sensitive customer information. Federated models can be trained across multiple institutions, allowing for more robust analysis while being in compliance with very strict data privacy regulations.[8]
- Other Privacy-Sensitive Domains: Federated analytics can be applied to any number of domains such as insurance, telecommunications, and public safety, where the primary concern is privacy. This could be shared for predictive maintenance, anomaly detection, and analysis of customer behavior without violating confidentiality.

3. Federated analytics has the potential to open doors towards cross-sector collaborations wherein organizations from multiple industries can share their insights, resources, and models while ensuring that the private data remains confidential. Supply chain management, smart cities, and energy management are sectors that would need integration of various data sources to make proper decisions.[9]

VII.CASE STUDIES AND REAL-WORLD APPLICATIONS

A. Examples of Federated Analytics in Use

Federated analytics is increasingly being implemented in sectors such as healthcare and finance, which are critical when data privacy has to be the top priority. These sectors frequently handle sensitive personal data, for which they enjoy the ability to collaborate and share insights without centralizing data.

1. Healthcare

- Medical Research and Diagnostics: Federated analytics has accelerated medical research since institutions can work together without transferring sensitive data from patients. For example, in medical imaging, hospitals and clinics can train deep learning models using various datasets of other institutions for the detection of tumors or for identifying rare diseases. Data aggregation can thus occur across different hospitals without sending over sensitive images that reduce the threat of privacy issues.[9]
- Discovery of drugs: Pharmaceutical companies have already applied federated learning for drug discovery. It is possible to pool insights coming from clinical trials and research studies by using decentralized models without requiring the various institutions to share sensitive patient records or proprietary data. A good

example would be the use of federated analytics for predictions of effectiveness of new drugs in combinations between several research centers while preserving privacy and intellectual property.

2. Finance

- Fraud Detection: Federated analytics is being applied by financial institutions to detect fraud across their network without revealing transaction data. For example, banks collaborate on a shared machine learning model that will identify suspicious behavior or fraud patterns, such as unusual transactions or identity theft. The federated model allows for the aggregation of insights from a large number of institutions, and this results in more robust fraud detection systems.[10]
- Federated learning: another application includes in finance federated learning toward improvement in models related to scoring and assessing credits. Here financial institutions share information while protecting critical information regarding its customers for achieving accurate credit worthiness-based models. Those diverse federated models so developed are highly competent for developing precision about whether or not it would be correct in forecasting probable defaults on the loans, or possibly risky investment.

B. Industry Adoption

The federated analytics implementation is already widespread among many influential companies and other institutions in industry sectors, combining both privacy-preserving and benefits of collaboration

- Google: For the tech firms, Google Gboard, one of the company's virtual keyboard app, is what has spearheaded the use of federated learning. Improvements of the model for the prediction texts and languages using federated learning are done without utilizing the necessity of forwarding a user's data to central servers of those firms. A firm trains their models in ML at users' devices through the cell phone. Accuracy is thereby achieved without transmitting the sensitive information. Examples of these may include private personal text or input. This is an excellent example of federated analytics at scale, where millions of devices are continuously working to train models in real-time without compromising on the privacy aspect.

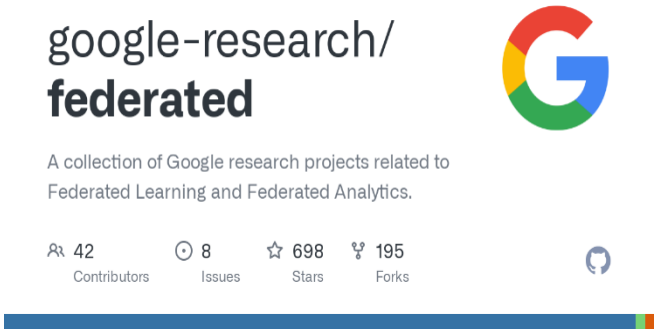

Fig. 5. Github Repo of Google research projects related to Federated Analytics

- Apple: Apple has used federated learning by integrating it into its ecosystem, especially in healthcare applications. It trains the health data, including heart rate monitoring, sleep tracking, and fitness metrics, on machine learning models directly on users' devices. That means that the user's personal health information never leaves his or her own device. Thus, it ensures that Apple would be able to develop better algorithms for health and fitness while completely respecting user's privacy. Another application of Federated Learning outside of health research is enhancing models for Siri-the predictive typing, voice recognition, models.
- Federated learning was the basis for several medical research projects that hospital networks and research institutions like Mayo Clinic, Cleveland Clinic, and Mount Sinai Health System conducted together. The research data could be aggregated by the hospital networks in a manner that they would help improve the predictability of the analytics that may be available for patient care, diagnostic tools, and treatment strategies strictly within and of the parameters set forth for adherence to privacy regulations such as HIPAA. It has been possible to collaborate with the development of AI models that detect early symptoms of diseases like Alzheimer's or cancer without the transfer of sensitive patient information.

- Financial Institutions: Global banks and finance houses, such as HSBC, JPMorgan Chase and Mastercard, are also using federated analytics for fraud detection, risk management and customer behavior analysis. These institutions have realized that federated learning improves the accuracy of predictive models, especially fraud detection, where insights from several banks will result in a better detection of unusual patterns in various types of transactions. In this regard, Deutsche Bank has applied federated learning to enhance its anti-money laundering systems to keep in line with regulations without breaching customer privacy.
- Telecommunications Industry: AT&T and Vodafone, among others, are exploring federated analytics to improve predictive maintenance, optimize network performance and predict churn from customers. Thus, they would update their models with insights drawn from different regions or network infrastructures without needing to centralize user data to preserve customer information while reaping the knowledge generated across their networks.

VIII. CONCLUSION

Federated analytics is a revolutionary way in which organizations can draw on data for insight without necessarily breaching privacy or compromising security. This is achieved through the ability to train models over decentralized datasets so that sensitive data need not be centralized and, therefore, avoid the possibility of a breach of privacy while simultaneously complying with strict industry requirements set forth by such sectors as healthcare and finance. It is a very powerful tool for industries dealing with huge amounts of sensitive information due to statistical analysis combined with machine learning algorithms and privacy-preserving technologies, such as differential privacy and secure aggregation.

Applications range from fraud detection in finance to better diagnostics in healthcare. Of course, a much more concrete validation of the importance of federated analytics is now coming from growing companies such as Google, Apple, and most of the major financial institutions adopting federated learning and showing promise with federated analytics. Such federated learning enabled cross-institutional collaborations lead to applications that not only increase the precision of machine learning models but could be impossible to achieve without federated learning.

Federated analytics, therefore, suffers from problems like scalability, convergence of a model, and homogeneity in data sources, even though it seems to be highly promising. In this regard, research and development desperately seek the emergence of improvement into the efficiency, security, and accuracy of federated analytics, which also offers avenues for wide adoption and success in all avenues of industrial sectors.

Therefore, federated analytics might very well occupy a key position in the near future handling both the imperative to collaborate at the level of insights with the mandate to protect data and hence mark a new beginning of an era-defining how the institutions across the lines of sector will handle the issue of information sharing followed by the process of its subsequent analysis into decision support.

REFERENCES:

1. Wang, D., Shi, S., Zhu, Y., & Han, Z. (2021). Federated analytics: Opportunities and challenges. *IEEE Network*, *36*(1), 151-158.
2. Elkordy, A. R., Ezzeldin, Y. H., Han, S., Sharma, S., He, C., Mehrotra, S., & Avestimehr, S. (2023). Federated analytics: A survey. *APSIPA Transactions on Signal and Information Processing*, *12*(1).
3. Pandey, S. R., Nguyen, M. N., Dang, T. N., Tran, N. H., Thar, K., Han, Z., & Hong, C. S. (2021). Edge-assisted democratized learning toward federated analytics. *IEEE Internet of Things Journal*, *9*(1), 572-588.
4. Chen, D., Wang, D., Zhu, Y., & Han, Z. (2021). Digital twin for federated analytics using a Bayesian approach. *IEEE Internet of Things Journal*, *8*(22), 16301-16312.
5. Wang, Z., Zhu, Y., Wang, D., & Han, Z. (2022, May). FedFPM: A unified federated analytics framework for collaborative frequent pattern mining. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications* (pp. 61-70). IEEE.

6. Parra-Ullauri, J. M., Zhang, X., Bravalheri, A., Moazzeni, S., Wu, Y., Nejabati, R., & Simeonidou, D. (2024). Federated Analytics for 6G Networks: Applications, Challenges, and Opportunities. *IEEE Network*.
7. Froelicher, D., Troncoso-Pastoriza, J. R., Raisaro, J. L., Cuendet, M. A., Sousa, J. S., Cho, H., ... & Hubaux, J. P. (2021). Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nature communications*, *12*(1), 5910.
8. Harth, N., Anagnostopoulos, C., Voegel, H. J., & Kolomvatsos, K. (2022). Local & Federated Learning at the network edge for efficient predictive analytics. *Future Generation Computer Systems*, *134*, 107-122.
9. Margolin, E., Newatia, K., Luo, T., Roth, E., & Haeberlen, A. (2023, October). Arboretum: A planner for large-scale federated analytics with differential privacy. In *Proceedings of the 29th Symposium on Operating Systems Principles* (pp. 451-465).
10. Yue, X., Kontar, R. A., & Gómez, A. M. E. (2024). Federated data analytics: A study on linear models. *IISE Transactions*, *56*(1), 16-28.
11. https://github.com/google-research/federated