**Data Privacy Preserving in Payment Transactions**

Subhasmita Maharana- [smaharan@kent.edu](mailto:smaharan@kent.edu)

Keerthi Akhila Pasam- [kpasam@kent.edu](mailto:kpasam@kent.edu)

Mukthasree Vengoti- [mvengoti@kent.edu](mailto:mvengoti@kent.edu)

**Abstract:**

Incredible amounts of data are being generated by various organizations like hospitals, banks, e-commerce, retail, and supply chain, etc. by virtue of digital technology. The voluminous data generated from the various sources can be processed and analyzed to support decision making. However, data analytics is prone to privacy violations. Although data analytics is useful in decision making, it will lead to serious privacy concerns. Hence privacy preserving data analytics became very important.

In Today’s digital landscape, tons of data are generated across different industries, which can help make decisions but also puts privacy at risk. While data analysis is useful, it also raises privacy concerns, especially for organizations like the European Central Bank handling payment transactions.

This Report highlights the importance of data privacy and advocates for effective privacy-enhancing techniques and risk management to balance data utility with privacy protection.

1. **Introduction**
   1. **Motivation:**

The financial industry is significantly dependent on the analysis of payment transaction data for several reasons, including risk management, fraud detection, and consumer behavior analysis. However, the employment of traditional data analysis methodologies carries a risk of disclosing sensitive consumer information due to growing concerns about user security and data privacy rules.

Financial institutions can examine payment trends for fraud identification without compromising the privacy of individual transaction information by utilizing privacy-preserving approaches.   
Aggregated payment data can be analyzed by regulatory agencies to track trends and ensure financial stability without revealing personal client information.

**2.** **Project Description**

**2.1 Brief Description:**

This project European Central Bank (ECB) explores, compiles and, keeps enormous databases of payments made inside the Eurozone. Strict data privacy procedures are necessary to protect people's privacy and adhere to legal obligations like the General Data Protection Regulation (GDPR) because these datasets contain sensitive financial information.

Conducting comprehensive analysis of all payment transactions, implementing data anonymization techniques, and, assessing the effectiveness of anonymization methods to derive valuable information and to protect individuals privacy while concluding the analysis.

**2.2 Challenges and Technical Contributions:**

Extracting valuable insights from payment transaction data, anonymizing the dataset sufficiently while ensuring anonymity and compliance with data privacy regulations to protect individuals privacy poses a significant challenge.

Techniques such as aggregation, generalization, or noise addiction may impact the quality of data. Also analyzing the complex financial structures and relationships while maintaining privacy requires expertise in both data privacy and domain knowledge.

**2.3 Contribution of Team members:**

All the team members in this project have contributed equally so we are able to complete this project report quickly and efficiently.

Subhasmita has worked on collecting data, performed Data preprocessing methods, contributed to Final Analysis report.

Keerthi worked on visualizing the datasets, feature importance analysis, and contributed towards presentation.

Muktha has performed modeling techniques, Risk Stratification, Risk Aggregation analysis on dataset and, contributed towards Final Analysis report.

1. **Background**

**3.1 Related Papers:**

Dwork, Cynthia, et al. "Differential privacy." Proceedings of the thirty-third ACM symposium on Theory of computing. ACM, 2006.

Machler, Martin, et al. "Learning from data: A short course." Springer Science & Business Media, 2011.

**3.2 Software Tools:**

Programming Language: Python (with libraries like Pandas, NumPy, and SciPy)

Data Analysis Tools: Google Colab Notebook, Excel sheet

**3.3 Hardware:**

13th Gen Intel(R) Core(TM) i7

RAM 16.0 GB

**3.4 Related Programming Skills:**

Python programming, Data analysis libraries (Pandas, NumPy, SciPy, Matplotlib), basic understanding of differential privacy algorithms.

1. **Problem Definition**

**4.1 Formal Definition:**

This section typically refers to a formal mathematical definition, which might not be relevant in this project.This project aims to analyze the ECB's Payments transactions dataset while ensuring data privacy preservation.

Understanding the structure and content of the dataset to identify key variables and areas of interest for analysis.

**4.2** **Challenges of Tackling the Problem:**

The problem at hand involves analyzing payment transaction data from the European Central Bank (ECB)'s dataset while prioritizing data privacy preservation. With the financial industry heavily reliant on such data for risk management, fraud detection, and consumer behavior analysis, there's a critical need to safeguard sensitive consumer information in compliance with evolving data privacy regulations.

Financial institutions and regulatory agencies can leverage privacy-preserving approaches to extract insights from aggregated payment data without compromising individual transaction privacy. However, ensuring the efficacy of anonymization techniques while maintaining data utility poses significant challenges.

Challenges and Technical Contributions: The primary challenge lies in effectively anonymizing the dataset to protect individuals' privacy while retaining its analytical value. Techniques such as aggregation, generalization, and noise addition must be carefully implemented to balance privacy preservation with analytical accuracy.

**5. The Proposed Techniques**

**5.1 Framework (Problem Settings):**

In our proposed framework, we address the challenge of analyzing payment transaction data while ensuring data privacy preservation. The framework encompasses privacy preserving techniques such as differential privacy mechanisms, data aggregation and secure multiparty computation. By integrating these techniques, we aim to balance the need for insightful analysis with the imperative of protecting sensitive financial information.

**5.2 Details of Major Techniques:**

**5.2.1 Data Pre-processing Analysis:**

* The dataset has been loaded into DataFrame and some analysis steps have been performed to understand the data which includes as below:
* Displaying the structure of DataFrame.
* Checking for missing values if any using **isnull().sum(),** and removing them by using imputation method.
* Displaying the column names, data types of columns, summary Statistics of the numerical variables in dataset.
* Removing unnecessary columns from dataset.
* Removed outliers from the dataset, if there are any by calculating the z-score value considering or adjusting the threshold values accordingly to data.
* After removing the missing values and outliers from dataset there are **265764 observations and 19 variables.**

**5.2.2 Feature Engineering (Creating interaction terms):**

* Interaction terms have been created to satisfy our modeling and problem prediction techniques.
* Transaction\_Location term has been created from other 2 variables TYP\_TRNSCTN and REF\_AREA.
* New features have also been derived from TYP\_TRNSCTN, OBS\_VALUE, Transaction\_Count, and Avg\_Transaction\_Amount.

**5.2.3 Modeling Techniques Used in Analysis:**

We have used 2 modeling techniques i.e., Linear Regression and Random Forest Regressor.

* Linear regression can be applied to the analysis of payment transactions to forecast different features including transaction amounts, fraud possibilities, or risk scores. However, it can be applied to aggregated or anonymized data to ensure individual transaction details are not exposed.
* By finding patterns and trends in transaction data, linear regression models can aid in risk management. For example, they can assist in locating transactions that stand out or strange spending patterns that might point to fraud.
* And Random Forest Regression is a technique that works similarly to Linear Regression to forecast transaction amounts, fraud probability, or risk scores. It is appropriate for high-dimensional transaction data since, in comparison to linear regression, it can handle bigger and more complicated datasets.
* However, while managing sensitive financial information, it's crucial to put extra safeguards in place such data anonymization, aggregation, or encryption procedures to assure data privacy preservation.

**5.2.4 Feature Importance Analysis:**

**Finding Sensitive Features:** Feature importance analysis assists in determining which aspects of payment transactions are the most sensitive and holds important user data, such as payment amounts, personal information, or shopping preferences.

**Risk Assessment and Mitigation:** A more accurate assessment of the privacy concerns associated with different forms of data is made possible by an understanding of the significance of distinct attributes. Organizations can successfully reduce the possibility of data breaches or unauthorized access to sensitive information by giving priority to protection measures for high-risk characteristics.

**Anomaly Detection:** Organizations can more effectively identify irregularities or questionable activity in financial transactions by knowing the significance of certain features.

**5.2.5 Risk Stratification:**

Below are some points why is, risk stratification is used in data privacy preservation in payment transactions.

According to the feature we get from feature importance analysis, we stratified the risk into levels (i.e., low, medium, and high).

**Resource Allocation:** The risk associated with different types of data varies. Organizations can more effectively deploy their resources by stratifying the risks associated with various types of data. They can apply more lax security measures to less important data while concentrating their efforts on safeguarding the most sensitive data items.

**Efficient Prioritization:** By using risk stratification, companies can arrange their efforts to protect privacy according to the gravity of the possible outcomes. Higher importance is given to protecting data pieces that, if breached, could result in major financial loss, harm to one's reputation, or penalties from regulatory bodies.

**User Trust:** Users' confidence in payment systems and financial organizations is increased when sensitive data is protected. Organizations can enhance customer loyalty and satisfaction by showcasing their dedication to protecting client information by giving priority to the security of high-risk data items.

**5.2.6 Risk Aggregation:**

To present a thorough picture of the entire risk landscape, risk aggregation is the process of merging distinct hazards connected to various data pieces, transactions, or processes in the context of data privacy preservation for payment transactions.

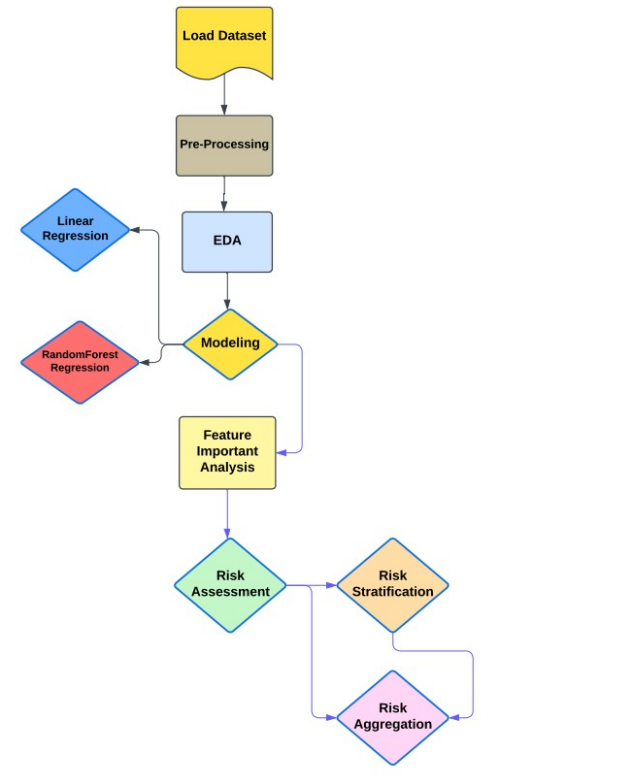
**Recognizing Cumulative Risks:** While individual data points or procedures might not be very risky, when compounded, they might lead to serious weaknesses. By identifying and assessing these cumulative risks, risk aggregation enables companies to handle not just individual threats but also their aggregate effects on data security and privacy.

**Prioritizing Mitigation Efforts:** Organizations can order their mitigation efforts according to the severity and possible consequences of a combination of vulnerabilities by aggregating risks. They can maximize the efficacy of their risk management solutions by concentrating efforts on high-risk areas where numerous factors converge.

**6. Visual Applications**

**6.1 Flow Chart:**

The below flow chart depicts a typical machine learning workflow for building a model. Each and every step is performed in this Final Analysis project on Data privacy preservation in Payment transactions:

****

* 1. **Visualizations**
     1. **Kernel Density Chart:**

This graph shows the distribution of transaction amounts in the dataset. Most transactions are smaller, but there are also a significant number of very high-value transactions, although the exact number is hard to determine due to the scale of the graph. Overall, it suggests a wide range of transaction amounts. This KDE (kernel density estimate plot) visualizes the median transaction amount is around 0.5.

It suggests that there is a wide range of transaction amounts in the payment transaction dataset, with a concentration of transactions around the median value. There are also a few outliers with very high or very low transaction amounts.

**A graph with numbers and symbols

Description automatically generated**

* + 1. **Correlation Heatmap:**

This heatmap visualizes the correlation between different variables in the dataset: number of transactions, number of decimals, and unit multiplier. The color intensity indicates the strength of the correlation, with deeper red showing a strong positive correlation and deeper blue showing a strong negative correlation. From the heatmap, we can see that there is practically no relationship between the number of transactions and the number of decimals. There also appears to be no correlation between the number of transactions or decimals and the unit multiplier. Remember, correlation doesn't mean causation!

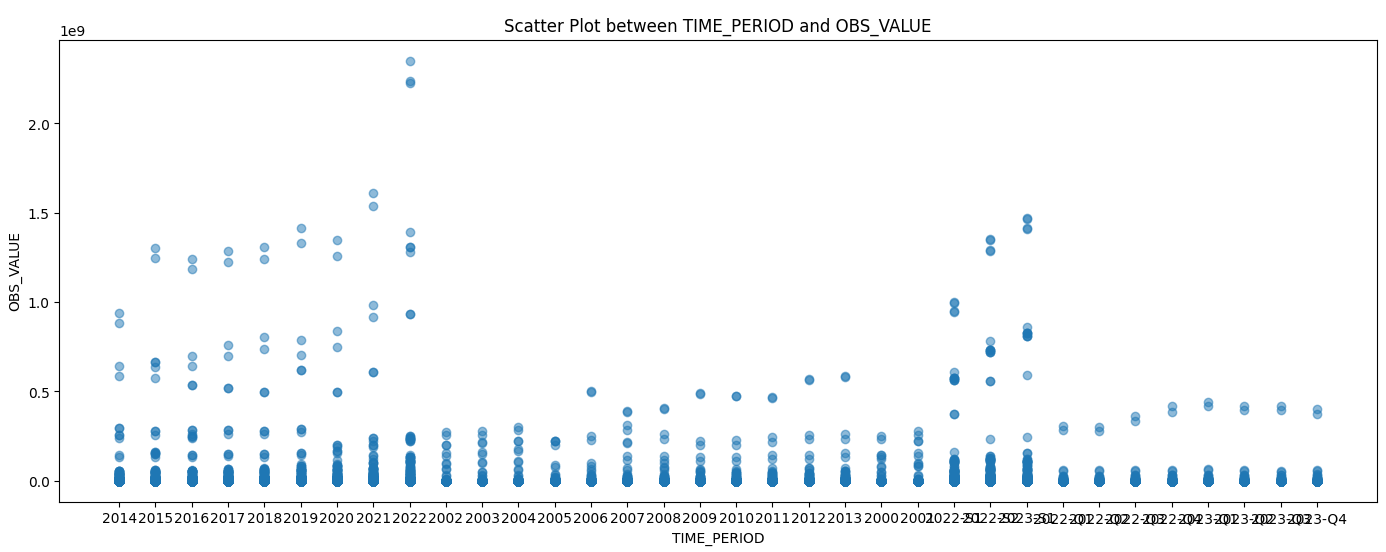
We can see that there is a weak negative correlation between the number of transactions (T) and the number of decimals (DECIMALS). The value is -0.01, which is very close to zero.

**A diagram of a heatmap

Description automatically generated**

* + 1. **Scatterplot visualization:**

This scatterplot helps us understand if there's any connection between the time when transactions happen and how much money is involved in those transactions. For example, we can see if transaction amounts tend to be higher or lower during certain time periods, like weekdays versus weekends.

****

**7. Experimental Evaluation**

**7.1 Experimental Settings:**

Dataset description: The dataset has been taken from the following website i.e., ECB Data Portal.

<https://data.ecb.europa.eu/data/datasets>

The above dataset describes about the various features which are related to European Central Bank portal consisting of KEY, FREQ, REF\_AREA, COUNT\_AREA, TYP\_TRNSCTN, RL\_TRNSCTN, FRD\_TYP, TRANSFORMATION, UNIT\_MEASURE, TIME\_PERIOD, OBS\_VALUE, OBS\_STATUS, CONF\_STATUS, TIME\_FORMAT, TIME\_PER\_COLLECT, DECIMALS, TITLE, TITLE\_COMPL, UNIT, UNIT\_MULT.

From these we have considered a few important features for modeling that are highly related to privacy according to the feature importance analysis.

**7.2 Evaluation Measures:**

* As explained in the above, we have done all the Data preprocessing, exploratory data analysis methods to understand the whole data and do the next steps.
* Features selected according to the problem statement:

Feature variables: except OBS\_VALUE, have taken all other variables

Target variables:OBS\_VALUE

**7.2.1 Linear Regression:**

We have performed the Linear Regression modeling technique to forecast different features including transaction amounts, fraud possibilities, or risk scores. Below are the evaluation measures obtained from Linear regression modeling:

A black text on a white background

Description automatically generated

**7.2.2 Random Forest Regression:**

We have also performed the Random Forest modeling technique that works similarly to Linear Regression to forecast transaction amounts, fraud probability, or risk scores. It is appropriate for high-dimensional transaction data since, in comparison to linear regression, it can handle bigger and more complicated datasets. Below are the evaluation measures obtained from Random Forest regression modeling:

A black text on a white background

Description automatically generated

From the above results of modeling techniques, we conclude that, considering the values of MSE, RMSE, and R-squared: the values obtained from linear regression are relatively high compared to Random Forest. MSE and RMSE values for Random Forest Regression are significantly lower than those for Linear Regression which indicates that the Random Forest model provides better predictions with small errors.

The R2 value for Linear Regression is close to zero, explaining little variance in target variable. And the R2 score for Random Forest Regression is (0.92), indicating that the Random Forest model is a better fit for the data.

Based on the results obtained, Random Forest Regression appears to be the more suitable model for predicting payment transaction values, offering better predictive performance, and capturing the underlying patterns in the data more effectively.

**7.2.3 Feature Importance Analysis:**

Here, we conduct the feature analysis, considering all the variables in the dataset, to calculate the privacy risk score.

**A screenshot of a computer code

Description automatically generated**

And from the results, we can interpret that Avg\_Transaction\_amount has 0.102, OBS\_STATUS has 0.115 indicating both have highest privacy risk value suggesting that they may contain more sensitive data. Title, completeness of the title has 0.092 value, indicating having the 2nd highest privacy risk value. COUNT\_AREA having 0.069 value and Transaction\_Location having 0.033 with moderate risk values. While features like KEY and TRANSFORMATION pose lower privacy risks. Additionally, features with negative scores like REF\_AREA and TIME\_PERIOD could potentially aid in reducing privacy risks through data anonymization or generalization.

A graph with blue and white bars

Description automatically generated

The graph shows how important each feature is in predicting the outcome of a machine learning model trained on the Payment Transaction Dataset. This plot here suggests that the KEY, reference area, count area, Time\_period, title, completeness of the title are the most important factors for the model to predict the outcome. While features such as Transaction\_count, Unit\_measure, Transformation, Transaction\_loaction, OBS\_STATUS are features with low importance.

**7.2.4 Risk Stratification:**

Performed Risk Stratification based on the features that were from feature importance analysis which provides valuable insights into the relative importance of different attributes prioritizing resources and efforts towards managing and mitigating high-risk factors.

**A screenshot of a computer program

Description automatically generated**

Risk values were stratified into three categories: low, medium, and high. Thresholds were defined to categorize the risk levels based on the magnitude of the risk values.

Each attribute was assigned a risk level based on its privacy risk value compared to the defined thresholds.

From the results, we can interpret that **COUNT\_AREA, TITLE\_COMPL, TITLE, OBS\_STATUS, Transaction\_Location and, Avg\_Transaction\_Amount** have been classified as having a **high-level privacy risk** and require closer attention. **REF\_AREA, TIME\_PERIOD and TRANSFORMATION** have **low risk levels** that aren’t a serious threat to privacy. While **KEY** has a **medium level of risk.**

**7.2.5 Distribution of Average Privacy Risk Levels:**

The plot visualizes privacy risk for various variables that were defined above according to the risk level values. And we see that the **COUNT\_AREA, TITLE\_COMPL, TITLE**, **Avg\_Transaction\_Amount**, **OBS\_STATUS,** **Transaction\_location** having high risk compared to other variables.

By the above result the banks can focus on strengthening controls and implementing measures to mitigate risks associated with high-risk attributes.

**A graph of a bar chart

Description automatically generated with medium confidence**

**7.2.6 Risk Aggregation:**

Risk aggregation refers to the process of combining individual privacy risk scores associated with various attributes of payment transactions into an overall risk score for each transaction.

The risk scores for individual attributes were aggregated using weighted averaging.

The aggregated risk score represents the comprehensive assessment of privacy risks associated with a particular transaction.

For our model the overall result according to feature analysis, risk stratification and risk aggregation is moderate.

An assessment of **Medium** overall risk suggests a balanced perspective on the privacy risks associated with payment transactions, where they are neither negligible nor overly significant, but rather fall within a moderate range.

**8. Conclusion:**

We have found potential privacy hazards connected to various payment transaction types by examining transaction variables and how they affect privacy risk scores. Based on the computed risk scores, we have estimated the amount of privacy risk connected to each transaction feature.

We have also examined the ways in which several transaction attributes, including location, kind, and others, influence the total risk to privacy. Understanding which attributes most influence observed transaction values and, in turn, privacy risk was made easier with the use of the feature importance analysis.

By allocating numerical scores for privacy risk, we have calculated the amount of privacy risk connected to every transaction attribute.

**9. Future Work**

**Customer Behavior Analysis:** Research on consumer activity patterns and how they affect privacy threats may prove to be a worthwhile endeavor. By combining the analysis of transactional data with customer demographics, interests, and interactions, risk concerns related to consumer groupings may become clear.

**Real-Time Monitoring & Alerting:** It might be investigated to develop real-time monitoring systems that can identify and address privacy problems in money transactions as they happen.

**Temporal Analysis:** Understanding how privacy risks in financial transactions change over time may be gained by performing a temporal analysis. Analyzing transaction data trends, seasonality, or cyclical patterns and their effects on privacy risk levels could be one way to do this.

**10. References**

[1] [M. John Basha](https://www.researchgate.net/scientific-contributions/M-John-Basha-2256136521?_sg%5B0%5D=X5A4RDnQIkw6q2D9m4JKPkxvR2bix3EeFQrIph-FLmqt7dyszXFqnhpeZMCsUQws0HW9jQM.zd6OljLaJR2BJNC3gRP3iNvyIbHBPJq3_8pMf7fyFLLFB5IEsHjJxKeyvT6PIGWcklBkVE6oX2NIR6g57IccwQ&_sg%5B1%5D=XqF91bMmfbuOsRIK0EdDeC7RBRV3FUTFvsDrJkHc-XQ63wIVVAauFE3gMOAGbdWVfRA1_uc.GByqNoDuNXZjL-CcYTYZRiOnRvuQdXfwWYsIi5avmoV9Pz5lSL5BuS4K3WF9cZlE0Aj6FxsKERsV3E_1oKwtRw&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [T. Satyanarayana Murthy](https://www.researchgate.net/scientific-contributions/T-Satyanarayana-Murthy-2253278281?_sg%5B0%5D=KbtD3GJ-IZnnDpg-Qe3fOJSfsD27obHvRvyC-Q9kjPfi-1mTdBzA7MJGGg_Q6_xzXRLPoQc.-9I3Fxl7Q61DQyt1idhfoPxtCI_bXIu2_LW9r23t5d9Xyx2-BV7Hy8fslIAONaiOu4tUZUJyPtK3cqoF7H8UJA&_sg%5B1%5D=0V2Fa_jLPR2IZiHY_Ml7sOkXW0rDn_ZACCkXKUsWEU93f9bQI9NkoY-PwbO94MdmqMjG7qQ.o9V1B0EN8iFYU9TDlP_BVnh6X-wx6MuUKZf3ZZL_ZdM66UVdmDaTr5dF6p3Oe_s2jTA-997r3pwkrtwC_PbUrw&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [A.S. Valarmathy](https://www.researchgate.net/scientific-contributions/AS-Valarmathy-2273615876?_sg%5B0%5D=KbtD3GJ-IZnnDpg-Qe3fOJSfsD27obHvRvyC-Q9kjPfi-1mTdBzA7MJGGg_Q6_xzXRLPoQc.-9I3Fxl7Q61DQyt1idhfoPxtCI_bXIu2_LW9r23t5d9Xyx2-BV7Hy8fslIAONaiOu4tUZUJyPtK3cqoF7H8UJA&_sg%5B1%5D=0V2Fa_jLPR2IZiHY_Ml7sOkXW0rDn_ZACCkXKUsWEU93f9bQI9NkoY-PwbO94MdmqMjG7qQ.o9V1B0EN8iFYU9TDlP_BVnh6X-wx6MuUKZf3ZZL_ZdM66UVdmDaTr5dF6p3Oe_s2jTA-997r3pwkrtwC_PbUrw&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [N. Parkunam](https://www.researchgate.net/scientific-contributions/N-Parkunam-2256139635?_sg%5B0%5D=KbtD3GJ-IZnnDpg-Qe3fOJSfsD27obHvRvyC-Q9kjPfi-1mTdBzA7MJGGg_Q6_xzXRLPoQc.-9I3Fxl7Q61DQyt1idhfoPxtCI_bXIu2_LW9r23t5d9Xyx2-BV7Hy8fslIAONaiOu4tUZUJyPtK3cqoF7H8UJA&_sg%5B1%5D=0V2Fa_jLPR2IZiHY_Ml7sOkXW0rDn_ZACCkXKUsWEU93f9bQI9NkoY-PwbO94MdmqMjG7qQ.o9V1B0EN8iFYU9TDlP_BVnh6X-wx6MuUKZf3ZZL_ZdM66UVdmDaTr5dF6p3Oe_s2jTA-997r3pwkrtwC_PbUrw&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Privacy-Preserving Data Mining and Analytics in Big Data.

[2] [Surapon Riyana](https://www.researchgate.net/profile/Surapon-Riyana?_sg%5B0%5D=g0kEcYY-F7Ha2Ct_pA2AlFlWqpwOhheZSUp6rcwxJlMBu5stfBKfR5S-Ufk8k-8DZ21A0ys.ucieeMMshGrjIRdOVjRqC9TCoD_6uyxTS9uzq5dadp-NfMJupXyq0cEH4rflqg3_quq8-KMo_lfmcOEaPZnvsg&_sg%5B1%5D=T9INGvNFGbaH8mxS0VCo07P8oG90qXBR9ewc4H6x9msaH_d3iSgAdU4ZZ9D4NVFEda2MTaE.tHKNp1zKGU2MBkAgRRHAUHTz7SCAhwXke0fGheY9l7sh0aEaaU_aak29itLtaytWzK45a6RB9O9PEguNlzvNoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [Kittikorn Sasujit](https://www.researchgate.net/profile/Kittikorn-Sasujit?_sg%5B0%5D=g0kEcYY-F7Ha2Ct_pA2AlFlWqpwOhheZSUp6rcwxJlMBu5stfBKfR5S-Ufk8k-8DZ21A0ys.ucieeMMshGrjIRdOVjRqC9TCoD_6uyxTS9uzq5dadp-NfMJupXyq0cEH4rflqg3_quq8-KMo_lfmcOEaPZnvsg&_sg%5B1%5D=T9INGvNFGbaH8mxS0VCo07P8oG90qXBR9ewc4H6x9msaH_d3iSgAdU4ZZ9D4NVFEda2MTaE.tHKNp1zKGU2MBkAgRRHAUHTz7SCAhwXke0fGheY9l7sh0aEaaU_aak29itLtaytWzK45a6RB9O9PEguNlzvNoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [Nigran Homdoung](https://www.researchgate.net/profile/Nigran-Homdoung?_sg%5B0%5D=g0kEcYY-F7Ha2Ct_pA2AlFlWqpwOhheZSUp6rcwxJlMBu5stfBKfR5S-Ufk8k-8DZ21A0ys.ucieeMMshGrjIRdOVjRqC9TCoD_6uyxTS9uzq5dadp-NfMJupXyq0cEH4rflqg3_quq8-KMo_lfmcOEaPZnvsg&_sg%5B1%5D=T9INGvNFGbaH8mxS0VCo07P8oG90qXBR9ewc4H6x9msaH_d3iSgAdU4ZZ9D4NVFEda2MTaE.tHKNp1zKGU2MBkAgRRHAUHTz7SCAhwXke0fGheY9l7sh0aEaaU_aak29itLtaytWzK45a6RB9O9PEguNlzvNoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Privacy-Enhancing Data Aggregation for Big Data Analytics.

[3] [Soumia Zohra El Mestari](mailto:soumia.elmestari@uni.lu), Gabriele Lenzini, Huseyin Demirci: Preserving data privacy in machine learning systems.

[4] [Surapon Riyana](https://www.researchgate.net/profile/Surapon-Riyana?_sg%5B0%5D=g0kEcYY-F7Ha2Ct_pA2AlFlWqpwOhheZSUp6rcwxJlMBu5stfBKfR5S-Ufk8k-8DZ21A0ys.ucieeMMshGrjIRdOVjRqC9TCoD_6uyxTS9uzq5dadp-NfMJupXyq0cEH4rflqg3_quq8-KMo_lfmcOEaPZnvsg&_sg%5B1%5D=T9INGvNFGbaH8mxS0VCo07P8oG90qXBR9ewc4H6x9msaH_d3iSgAdU4ZZ9D4NVFEda2MTaE.tHKNp1zKGU2MBkAgRRHAUHTz7SCAhwXke0fGheY9l7sh0aEaaU_aak29itLtaytWzK45a6RB9O9PEguNlzvNoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [Kittikorn Sasujit](https://www.researchgate.net/profile/Kittikorn-Sasujit?_sg%5B0%5D=g0kEcYY-F7Ha2Ct_pA2AlFlWqpwOhheZSUp6rcwxJlMBu5stfBKfR5S-Ufk8k-8DZ21A0ys.ucieeMMshGrjIRdOVjRqC9TCoD_6uyxTS9uzq5dadp-NfMJupXyq0cEH4rflqg3_quq8-KMo_lfmcOEaPZnvsg&_sg%5B1%5D=T9INGvNFGbaH8mxS0VCo07P8oG90qXBR9ewc4H6x9msaH_d3iSgAdU4ZZ9D4NVFEda2MTaE.tHKNp1zKGU2MBkAgRRHAUHTz7SCAhwXke0fGheY9l7sh0aEaaU_aak29itLtaytWzK45a6RB9O9PEguNlzvNoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [Nigran Homdoung](https://www.researchgate.net/profile/Nigran-Homdoung?_sg%5B0%5D=g0kEcYY-F7Ha2Ct_pA2AlFlWqpwOhheZSUp6rcwxJlMBu5stfBKfR5S-Ufk8k-8DZ21A0ys.ucieeMMshGrjIRdOVjRqC9TCoD_6uyxTS9uzq5dadp-NfMJupXyq0cEH4rflqg3_quq8-KMo_lfmcOEaPZnvsg&_sg%5B1%5D=T9INGvNFGbaH8mxS0VCo07P8oG90qXBR9ewc4H6x9msaH_d3iSgAdU4ZZ9D4NVFEda2MTaE.tHKNp1zKGU2MBkAgRRHAUHTz7SCAhwXke0fGheY9l7sh0aEaaU_aak29itLtaytWzK45a6RB9O9PEguNlzvNoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Aggregate Query Frameworks for Preserving Privacy Data in Big Data Analytics.

[5] [Alaa Alwabel](https://www.researchgate.net/profile/Alaa-Alwabel-2?_sg%5B0%5D=CYEc3QrnJUwsGv87yFoC00Pu_yc57nDt5Qu-GlZi0WIYEvG-vQlSxghl8tc_pRUWjwD78wI.6yXvDWGUHXq1od0mYMhdbPwjXY4UHS-SmOq17wctTm0_TQbaGv1cFaL_5RwD9wt7gzw-PG6D-Q45i0GPPHrQnA&_sg%5B1%5D=6P6lHhHzyhQDeV6tmuT7bdzxEbqTjLbNpoD8Lj8ri1Ob3hJkhMXecQiOw7UqiZ8HSKTGGw0.GG_kkpuEb3N_GnTiKokfxiCdNSk3ceXcesodXrr3N1bpFQvMFJSGb1bXTRmCkYeHx0lSN0k144Ghh4bA-rmA7g&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Privacy Issues in Big Data from Collection to Use.

[6] [Mohammad Raeini](https://www.researchgate.net/profile/Mohammad-Raeini?_sg%5B0%5D=bBLzzjvnM-21KRBdtbjufviFbCQE-vF9eUH0uHNkfrcUvphzSEj_jTch9dY571MKbcIXFDI.7wutz7k-1ZA96LPNYi0OVX3HkUY2-nXrgObsNFSl2O-9PXqWMrr2ujJH2gHERbXRL5tlwKzUkFnij8E2uD-dRg&_sg%5B1%5D=BGO_PQGlTdGElcxMQl0B6XE7hD2KP4mQZ7LuxcjxiN9vc1ESvWk4U5xv3VH0_mdlaj7eGwk.DyGDDp9c1gfkiWaFVQrrZzoYSme1oeDSleMJwkGgaByvJywTSawIj9EP6msaNy8tXzIVOa43yJV_BLyOZJyxFw&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Privacy-Preserving Big Data Analytics: From Theory to Practice.

[7] [Houda Ferradi](https://www.researchgate.net/profile/Houda-Ferradi?_sg%5B0%5D=jY8Aevzxq8tHZ71UmuD7ARL88FvY1MxMXquJaHlSp_h_tDEMTZcE2dpDIuUBsqFnU46B3Js.VHsRhs_ESC9lW6dgc43jV3tJPnx7N4XUKsBJdIyfAbsrKz8ytI0wTzH4iBWfDwVG3WP_fgO8hQoZygCq4UWdxA&_sg%5B1%5D=vKqECOWe2VKu9_bcLErbfV7KhvcrV3JOaqEYLkzp2IJ5GLSwmKM9Emg6obOCHoxzBZP5PWg._MKhuYoaOEql0vxxNvczESC6QB-3AODhg8Z4BSBtOFivEj---JIpQyWEbpM_kn00d9uYs418EaSn22lZhmhuoQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Security and Privacy in Big Data Sharing: State-of-the-Art and Research Directions.

[8] [Arun Amaithi Rajan](https://www.researchgate.net/profile/Arun-Amaithi-Rajan?_sg%5B0%5D=40gfnt4rIvem_JX0ECVm2RoghgNcSiZuAQ5V5ABYtJ40zFSxjR638a3Rsc8ITvp-mDsfj7A.Z5H0nM7_yr622YLXqHeV2JdFMoSOm863TjnR6n6unGZTaELJjHFfrvZgfd0gGwnsF5amXX6VJ1Kjb6QXjO7n-w&_sg%5B1%5D=m4_NrSyxEaEKdaP28lFwwYk9jgpd7foqRO0pT6n9mDKLQVx3N96yBxrtXqyQqRZWUDmLNJs.EOoQ5RqC-0KesX7O3SFpxSCcx0NyqtHXUNbojaJqTSKlIhxywrgX0K6MOrbdL7lolXDQzqipswcVZ9E3I4mHYQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [V. Vetriselvi](https://www.researchgate.net/profile/V-Vetriselvi-2?_sg%5B0%5D=40gfnt4rIvem_JX0ECVm2RoghgNcSiZuAQ5V5ABYtJ40zFSxjR638a3Rsc8ITvp-mDsfj7A.Z5H0nM7_yr622YLXqHeV2JdFMoSOm863TjnR6n6unGZTaELJjHFfrvZgfd0gGwnsF5amXX6VJ1Kjb6QXjO7n-w&_sg%5B1%5D=m4_NrSyxEaEKdaP28lFwwYk9jgpd7foqRO0pT6n9mDKLQVx3N96yBxrtXqyQqRZWUDmLNJs.EOoQ5RqC-0KesX7O3SFpxSCcx0NyqtHXUNbojaJqTSKlIhxywrgX0K6MOrbdL7lolXDQzqipswcVZ9E3I4mHYQ&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19): Systematic Survey: Secure and Privacy-Preserving Big Data Analytics in Cloud.

**Additional Links to our Presentation and Demo:**

Google Colab Link: <https://colab.research.google.com/drive/1h2iEsEdQxtP47ORpaLQ-Cslj5H3ikPSF?usp=sharing#scrollTo=nc8LuoRThcG8>

Presentation Link:  <https://video.kent.edu/media/Project_Presentation_Group41/1_1nt4fn5r>

Presentation Demo: <https://video.kent.edu/media/Project_demo+12A14A11+pm/1_gjdrk97i>