Privacy Preserving and Information Sharing in Agriculture

Hadeel Almaimani, Erman Ayday, Rajbabu Velmurugan, Maryam Shojaei

hfa12@case.edu, exa208@case.ed, rajbabu@ee.iitb.ac.in, mshojaei@ee.iitb.ac.in

Introduction

- Farmers are left on their own to defend their privacy. At this point, there are no regulations that directly protect farmer's data.
- Some farmers may opt not to participate in data communities for fear that others may unfairly benefit from their data as a result they lose the opportunity to participate in negotiations with landowners, retailers, and other service providers.

Privacy Risks

Figure 1: Privacy Risk Graph

- Production data may be used by agriculture service providers not only to benefit producers by providing managerial decision support, but also to price discriminate.
- Data used to speculate in commodities markets with information that is not knowable to market participants raises concern about market manipulation.
- Unauthorized disclosure of soil, crop, and agriculture purchase information can cause severe economic losses to individual farmers.

Dataset

- Data obtained from Kaggle was used to characterize four farm units and their associated crops:

	Nitrogen	Phosphorous	Potassium	Temperature	Humidity	pH	Rainfall
Farm 1 (200)	22	133	200	23.24	87.10	6.0	91.13
Farm 2 (517)	100	42	39	26.05	80.84	6.4	70.96
Farm 3 (474)	62	42	35	26.51	77.56	6.4	190.14
Farm 4 (1009)	26	49	29	25.44	60.74	6.6	81.84

Table1: Dataset Attributes Aggregated by Machine Learning

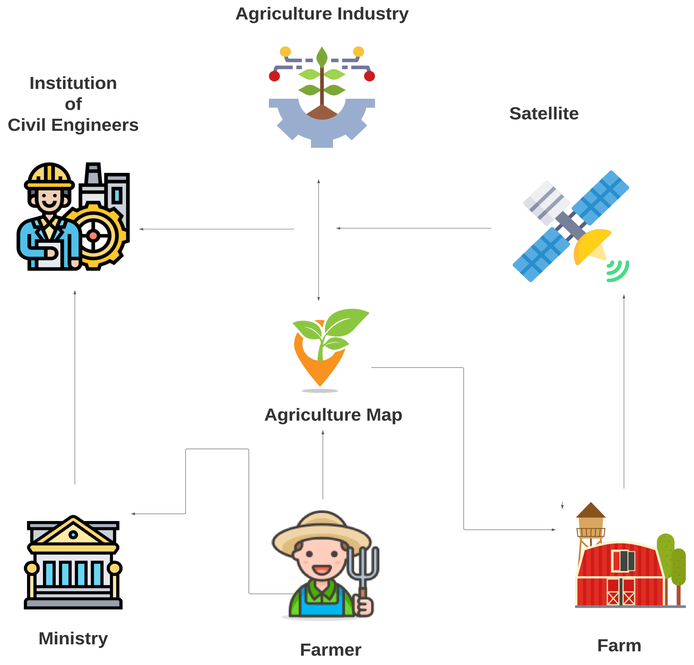

Figure 2: Continuity Scenario Affects on Privacy

Analysis

- The analysis focuses on epsilon-differential level of privacy that farmers may rely upon when sharing information about the attributes of their farm lots.
- In order to identify potential farm attributes, an analysis of the "Crop Recommendation Dataset" from Kaggle was performed using a k-nearest neighbors (KNN) algorithm.
- Cluster centers were used to group the various rows of the dataset into farm units. Using these farm units, data about which crops were associated with each cluster was extracted.
- Our goal is to evaluate the effect of adding a data element to the set.
- Additionally, we aim to contrast the changes in the machine learning model to the resulting analysis and make claims about the effect on the privacy of the individual farmers.

References

- Coble K., Mishra A., Ferrell S., and Griffin T. (2017). *Big Data in Agriculture: A Challenge for the Future. Applied Economic Perspectives and Policy*, (40)1, 79-96.
- Gupta M., Abdelsalam M., Khorsandroo S., and Mittal S. (2020). *Security and Privacy in Smart Farming: Challenges and Opportunities.* IEEE Access, Vol. 8, 34564-84. http://doi.org/10.1109/ACCESS.2020.2975142
- Ingle, A. *Crop Recommendation Dataset.* 2021. Kaggle. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset
- Li, M., Chow, S., Hu, S., Yan, Y., Shen, C., and Wang, Q. *Optimizing Privacy-Preserving Outsourced Convolutional Neural Network Predictions.* 2020. https://arxiv.org/pdf/2002.10944.pdf
- Scikit-Learn Developers. *sklearn.cluster.KMeans.* 2023. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

Preliminary Results

- The pie charts produced during analysis distribute the crops among four farms
- The original dataset after being grouped into four farms:

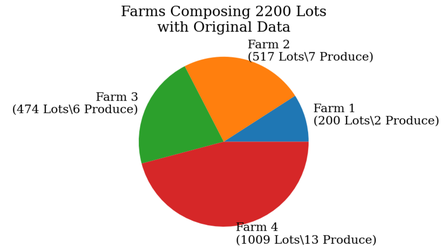

Figure 3: Crop Distribution Prediction on Original Dataset

- This figure demonstrates the effects of adding a manufactured element to the original dataset:

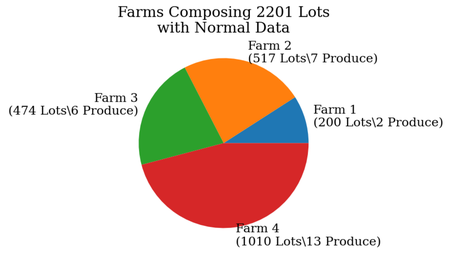

Figure 4: Crop Distribution Prediction on Modified Dataset

Conclusion

- Sophisticated analysis of the information that farmers provide can enable potential attackers with the information necessary to persuade investors.
- The primary technique for securing the privacy of farmers that is offered by adding 'noise' to the publicly available information and relying on epsilon-differential privacy.
- Ongoing, the dataset can be converted to time series data and investigation of the farmer-insurance relationship may be pursued.