# HW1: Machine Learning with Encrypted Data Annotated Bibliography

Brandon Hosley

June 14, 2020

[1] M. Barni, P. Failla, V. Kolesnikov, R. Lazzeretti, A. R. Sadeghi, and T. Schneider, "Secure evaluation of private linear branching programs with medical applications," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5789 LNCS, no. July 2015, pp. 424–439, 2009.

> This paper describes a method for employing branching programs for classification on encrypted data. They then propose how this method may be used in a way that allows a data-owning client (such as a health-care entity) to use the service of an analysis provider without either entity having to expose their sensitive or propriety data to the other.

[2] R. Bost, R. A. Popa, S. Tu, and S. Goldwasser, "Machine Learning Classification over Encrypted Data," in *NDSS*, vol. 4324, 2015, p. 14.

> The researchers create 3 different types of classifiers for training machine learning models with encrypted data. They produce a library with tools for constructing classifiers similar to the ones demonstrated in their work. Additionally, they provide information about the efficiency of their classifiers when applied to different Machine-Learning algorithms.

[3] C. z. Gao, Q. Cheng, P. He, W. Susilo, and J. Li, "Privacy-preserving Naive Bayes classifiers secure against the substitution-then-comparison attack," *Information Sciences*, vol. 444, pp. 72–88, 2018. [Online]. Available: https://doi.org/10.1016/j.ins.2018.02.058

> In this article, researchers propose a method of employing machine learning using a Naive Bayes classifier and additively homomorphic encryption. The result is a model that can be trained on encrypted data, allowing privacy on both the training and database sides. A primary vulnerability that the researchers are concerned with addressing is the Substitution-Then-Comparison type of attack.

[4] C. Gentry, "A Fully Homomorphic Encryption Scheme," *Dissertation*, no. September, p. 169, 2009. [Online]. Available: http://cs.au.dk/ stm/local-cache/gentry-thesis.pdf

This is Dr. Craig Gentry's PhD dissertation in which he proposes what is regarded as the first fully homomorphic encryption model. He describes at length all of the constituent aspects of his model, which includes significant discourse about partial homomorphic schemas and the ideal lattices that his model leverages. Dr. Gentry also proposes potential applications of this model.

[5] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in *Annual International Cryptology Conference*. Springer, 2000, pp. 36–54.

Researchers provide a proposal for a method of Data Mining multiple sources while preserving privacy for those data sources. In their example, privacy is preserved by having data owners perform necessary calculations on their own data and the researcher effectively collating their results. The suggested use for this method is to data mine across multiple health care sources without compromising the confidentiality of the data.

[6] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *International conference on the theory and applications of cryptographic techniques*. Springer, 1999, pp. 223–238.

Dr. Paillier investigates Composite Residuosity and a new 'trapdoor mechanism'. He uses this mechanism to propose three new encryption schemes. Two of the proposed schemes are additively homomorphic, and referenced in numerous other works in this bibliography.

[7] Y. Yasumura, Y. Ishimaki, and H. Yamana, "Secure naïve bayes classification protocol over encrypted data using fully homomorphic encryption," *ACM International Conference Proceeding Series*, 2019.

In this article the researchers examine training a machine learning model using a naive-Bayes classifier over data that has been encrypted with a fully homomorphic encryption algorithm. They additionally describe an optimization strategy that reduces the training time by 33% under the time taken by their initial implementation.