Towards secure digital farming: security model and risks associated to machine learning

A.B. Diallo[1], *S. Gambs*[2], *M.O. Killijia*[2] *and H. Lardé*[1]

[1]*Université du Québec À Montréal, Laboratoire de Bio-Informatique, P.O. Box 8888, Station Centre-Ville, Montréal, Québec H3C 3P8, Canada*
[2]*Université du Québec À Montréal, LATECE, P.O. Box 8888, Station Centre-Ville, Montréal, Québec H3C 3P8, Canada*
Corresponding Author: diallo.abdoulaye@uqam.ca

Abstract

In digital farming, machine learning is already widely used to optimize the production using sources such as genomics, health, welfare, production, and environmental data. However, this increasing use of machine learning has led to the emergence of multiple types of confidentiality and integrity breaches targeting both the models and the data they have been trained on. Our main objective in this paper is to discuss possible security issues that could arise in digital farming due to the use of machine learning techniques and the urgency to implement innovative countermeasures to prevent them. First, we propose a security model dedicated to the specific settings and threats of the digital farming context. In this model, we identify the resources at risk, define the different classes of actors, determine the risk vectors, and propose some realistic attack scenarios. Afterwards, we use this model to put in perspective the machine learning induced risks and show how they may adversely affect digital farming. The considered attacks encompass model theft, model inversion, membership inference, data poisoning and adversarial examples. For each of these threats, we also briefly revied possible mitigation means, such as differential privacy, prediction access control and robust statistics.

Keywords: Digital farming, cyber security, security model, machine learning.

Introduction

German Agricultural Society defines digital farming as the evolution of smart farming to better emphasize that nearly all aspects of farming now heavily rely on digital means (DLG Committee for Digitization, Work Management and Process Technology *et al.*, 2018). Collecting data massively from a wide variety of sources has allowed to take smart farming to a new level, leveraging big data to further improve the power of the decision-making system. To do so, various kind of data are collected such as environmental, production, health, welfare, genomics, and management. Machine learning (ML) is the core concept behind decision-making system, in which a sample of data called training dataset is used to generate a predictive model. ML is widely used across many industries and as ML techniques become, cybersecurity threats emerge (Papernot *et al.*, 2016) putting data confidentiality and production system integrity at risk.

Many actors have stated that cybersecurity is a concern in agriculture. For instance, a 2019 report from the U.S. Government Accountability Office (Dodaro, 2019), has indicated that improving cyber security should be one of the main priorities for actors in the agriculture sector. Geil *et al.* (2018) presented a survey in which they show that farmers are being directly affected at large scale. Window (2019) has conducted a study that presents major issues concerning data privacy, data ownership and level of attention given to cyber security in agriculture and all those points are requirements provided by the German Agricultural Society as well in their recent position paper.

A literature review allows us to conclude that working specifically on ML induced cybersecurity risks is a missing gap in the literature. Indeed, several reports focus on networking and Internet of Things (IoT) related risks (Gupta *et al.*, 2020; West, 2018) and several others on Big Data (Sykuta, 2016; Wolfert *et al.*, 2017). However, only a few works have been produced to study risks introduced by data analysis techniques in digital farming particularly in dairy industries. The U.S. Department of Homeland Security (Champion *et al.*, 2018) has also released a report in which they mention machine learning and at the regulatory level, farm data ownership is often present in the specialized literature (Sykuta, 2016; Window, 2019; DLG Committee for Digitization, Work Management and Process Technology *et al.*, 2018).

To study ML induced threats to digital farming, we first propose a security model adapted to this context of dairy farms, particularly in Canada, before studying the data life cycle and its interactions with the different resources and actors. Secondly, we propose an adversarial model to determine realistic threat vectors to ML systems in digital farming before proceeding with the investigation of the risks associated to ML, looking at five known vulnerabilities of ML systems and three possible practical mitigation strategies. Finally, we discuss another ML related security topics that should be investigated along ML induced threats.

Security model

Data chain: resources and actors

ML applications rely mainly on two assets: the training dataset and the learnt model. We will use the CIA framework to understand the impact of potential compromises on these resources. Confidentiality of the training dataset may be critical for privacy reasons, as for example valuable data such as genomics are used in digital farming, but also because it is part of the intelligence developed by ML application developers. Integrity of the training dataset is key to build reliable ML model and availability seems to only be a concern at the operational level. In addition, the confidentiality of the ML model is important in settings in which it is a monetizable resource such as ML as a Service (MLaaS), which is a form of pay-per-request service that could be compromised if the ML model was to be stolen. Furthermore, the ML model is a statistical representation of the training dataset as its confidentiality impact directly the confidentiality of the dataset. The integrity of the ML model is a concern for situations in which the ML predictions are used in a sensitive context such as farm management. Availability is a concern in time sensitive settings and for systems that cannot be substituted.

Wolfert *et al.* (2017) proposed a data chain in their framework for big data in smart farming. We adapted it to exclude network/infrastructure-based risk and focus on machine learning induced threats to assets described earlier. The principal node is data processing in which the ML model is developed. This node has two interfaces, the upstream data acquisition node in which the training dataset is being constituted and the downstream marketing node in which the end user is presented with a tool to query the model and obtain the associated predictions.

In Figure 1, we annotate the data chain to integrate actors found at each stage of the data lifecycle. The data provider is the main actor at the data acquisition stage, which is often the farmer but could also be laboratories in some cases (*e.g.* sample analysis on milk). We refer to the data collector to formalize the intermediary step consisting in

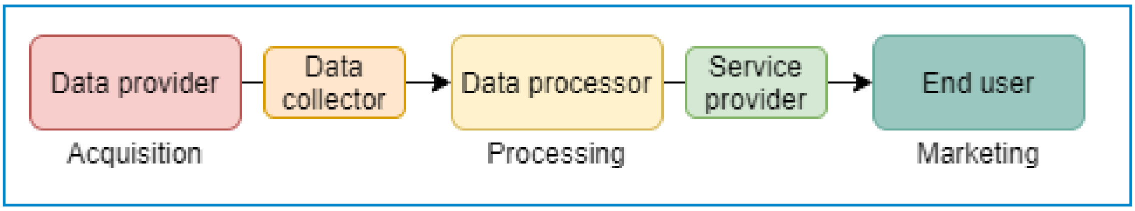

Figure 1. Actors involved in the data chain.

centralizing and storing the data. The data processor is the actor found at processing stage, who pre-process data to form the training dataset and train the model. Once the model is trained, the service provider refers to the actor in charge of wrapping the model in a product that can be distributed. Finally, the end user is the actor present at marketing stage that makes requests and applies predictions using the model through the marketed tool.

Adversarial model and threat vectors

We adopt the point of view of the data processor who forms the training dataset (preprocessing) and engineers the ML model (processing) because he is the actor having full and direct access to the resources that we aim to protect. We assume that the attacker could be or could impersonate a data provider (upstream) or an end user (downstream). For instance, in the case of Machine Learning as a Service (MLaaS), the attacker could target the model confidentiality for financial gain. In addition, if training dataset contains valuable information (e.g., genomics and/or production data), the attacker could target training dataset confidentiality. Finally, when the ML system is used for critical applications (food supply or seed production), the attacker could target integrity/availability of training dataset, ML model or prediction.

For the specific context of dairy digital farming, there can be a wide range of adversaries (insiders to the context of digital farming or not), thus leading to various levels of risks, ranging from a farmer seeking financial gain to eco-terrorists aiming to disrupt the food supply. As a result, the adversary is likely to have detailed knowledge about the digital farming and could have weak to strong technological skills. To study the attack surface and related threat vectors, we look at the interfaces of the ML system leaving aside all security concerns that are not inherently tied to ML (network, access control…).

Looking at the data chain, we have the upstream interface in which data is collected and preprocessed to form the training dataset and the downstream interface in which the trained model generates predictions upon user requests. At data collection stage, an attacker can craft and provide malicious data points to compromise the ML model and its predictions. At model interaction stage, malicious requests can lead to leak the ML model and the training dataset. More precisely, in controlled access settings, the attacker will have to compromise the model through a distant API whereas in model sharing settings the attacker is free to access both the program containing the model and the API locally, thus making the prediction access control harder and raising new concerns like reverse engineering.

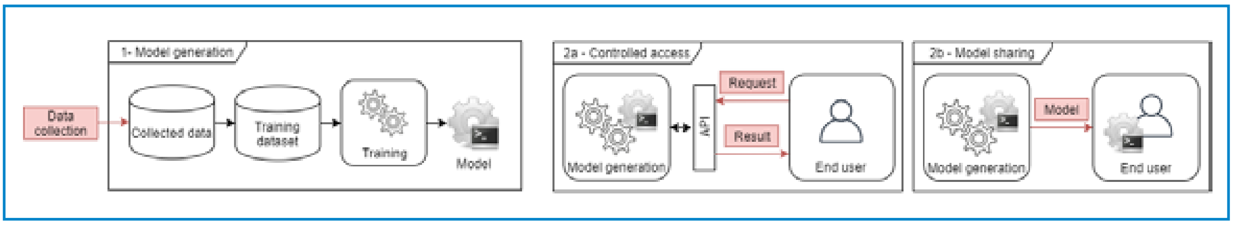

Figure 2. Attack vectors - (1) Upstream, (2) Downstream.

Risks to machine learning

Confidentiality of data and model

Membership inference attacks aim to deduce if a given data point is present in the training dataset or not. The first membership attack against a ML system was realized by Shokri *et al.* (2017). They targeted black box models in a context of MLaaS and were able to differentiate member data points only by sending requests to the model. Salem *et al.* (2018) have built upon this work to relax assumptions and extend the attack scenario. Others have studied membership on Generative Adversarial Network (Hayes *et al.*, 2019) and even on robust Deep learning techniques (Song *et al.*, 2019).

Model inversion attack aims to reconstruct information about a data point present in the training dataset. This type of vulnerability was introduced by Fredrikson *et al.* (2014) and demonstrated for a ML system used for personalized warfarin dosing. They were able to show that an attacker possessing the ML model and demographic information about a patient would be able to infer their genetic markers. Until recently, attacks failed to inverse higher complexity model such as neural networks but Y. Zhang *et al.* (2020) presented a technique that uses a small amount of auxiliary knowledge against neural network in white-box settings.

Finally, model theft attack aims to gain knowledge about a black box or grey box model such as type of algorithm, hyper-parameters or trained model parameters. Tramer *et al.* (2016) shows that model parameters at risk in a context of MLaaS, even when attacker does not have knowledge about the training data set or model algorithm. They use equation-solving attack to extract highly accurate model with a little number of requests. In their work presented earlier for membership inference Shokri *et al.* (2017) actually use a shadow model (*i.e.*, a model mimicking the target model behavior) as a step to mount their attack, essentially stealing a black-box version of the model.

Integrity of model and prediction

Data poisoning attack enables an adversary to influence the predictive power of a model by injecting malicious data points into the training data set. For example, Chen *et al.* (2017) present a scenario in which a back door is installed on a deep learning authentication system. Recently, the particular case of sequentially generated data for continuous learning system has been studied by X. Zhang *et al.* (2019). Finally, adversarial learning is a class of attacks in which an attacker exploits a predictive system by finding an input that induces an abnormal behavior of the system. For example, Al-Dujaili *et al.* (2018) successfully crafted adversarial example on malware binaries that allowed them to evade detection systems. Papernot *et al.* (2017) showed that they can instantiate such an attack in black-box settings targeting MLaaS.

Practical means of mitigation

Differential privacy is a designed privacy model to share information about a computation made on a dataset without compromising the privacy of each unique element. A possible implementation of differential privacy is through the addition of noise to the result to render unnoticeable the presence or absence of a particular profile in the data set. Differential privacy is a defense technique that is often used to counter membership inference and model inversion attacks. Since it can help generalizing the model, it can also be used to mitigate adversarial example and data poisoning attacks. Several attacks that we described take advantage of the accuracy of the predictions or use confidence levels shared along with the prediction. Controlling how the end user accesses the results of the predictive system (*e.g.*, by removing confidence levels) can help mitigating attacks such as membership inference and model inversion. Finally, robust models are designed to be more resilient to data perturbation both at training and prediction stage, helping mitigate adversarial crafting of data both at training stage (data poisoning) and at prediction stage (adversarial example). Several techniques can be employed to increase model robustness such as robust statistics, which are a class of estimation techniques that can be used to minimize impact of high diversity in statistical data distribution (*e.g.* outliers or small subgroups). Adversarial training is another technique in which adversarial examples are purposefully crafted and inserted in the training dataset to increase robustness against adversarial examples.

Discussion

In this paper, we have focused exclusively on threat vectors present at the data collection and the prediction interfaces. However, the security and privacy of ML systems security and privacy are also impacted by other concerns, which we briefly review in this section. In the data chain, in most cases the data collector (mostly the farmers) hand their data to data-processors. In addition, some of the data collected may be critical (*e.g.*, genomics) and considered as business secret (*e.g.*, production data). For this reason, the data collector might be reluctant to share its data which would break the first link of the data chain. In this scenario, homomorphic encryption is a cryptographic technique that could allow the data processor to train the ML system without the need for the data collector to divulge its valuable data, effectively maintaining our data chain functional. Data collectors have formulated concerns about the privacy of their data, and lack of cooperation between actors of the data chain have led to tensions within digital farming ecosystem. Doing our research, we have found multiple threat scenarios where the adversary is an insider to the digital farming context. Releasing tensions between actor would thus in itself help mitigate all ML induced risk to digital farming by lowering the likelihood of scenarios where the adversary is part of the data chain. A data trusts is a regulatory tool (a contract) designed to ensure that the management of a resource benefits each shareholder such as the resource provider (*i.e.* data-collector), resource processing agent (*i.e.* data-processor) and resource beneficiary (*i.e.* end-user).

Conclusion

As mentioned by organizational actors and looking at the context, it appears that security is a very concerning topic for digital farming that has been overlooked until now. We evaluated that ML was the left aside in terms of security and data privacy and dedicated our effort to help raise the attention as it is being used extensively. We have designed a Security model that helps framing the problem and investigated technical vulnerabilities and practical ways to mitigate them. During our study we also have found that related concern such as data ownership are directly impacting ML security and data privacy. Technical and regulatory tools such as homomorphic encryption and data trusts are available to help with these related concerns, effectively helping mitigating

ML induced risks. Agriculture has always been a technophile ecosystem and up to this day it has taken the most out of available technologies, leading to the digitalization of farming. We believe that digital farming sector should learn from other industries and take the opportunity to be ahead of the curve on security and data privacy concerns.

List of references

Abdullah Al-Dujail, Alex Huang, Erik Hemberg and Una-May OReilly, 2018. Adversarial deep learning for robust detection of binary encoded malware. 2018 IEEE Security and Privacy Workshops (SPW): 76-82.

Scott Champion, Linsky, Peter Mutschler, Brian Ulicny, Thomson Reuters, Larry Barrett, Glenn Bethel, Michael Matson, Thomas Strang, Kellyn Ramsdell and Susan Koehler, 2018. Threats to precision agriculture. U.S. Department of Homeland Security.

Xinyun Chen, Chang Liu, Bo Li, Kimberly Lu and Dawn Song, 2017. Targeted backdoor attacks on deep learning systems using data poisoning.

DLG Committee for Digitization, Work Management and Process Technology, DLG

Committee for Technology in Crop Production, DLG Committee for Technology in Animal Production, DLG Work Group Information Technology, Hans W. Griepentrog, Norbert Uppenkamp and Roland Hörner, 2018. Digital agriculture: A DLG position paper.

Gene L. Dodaro, 2020. Priority open recommendations: U.S. department of agriculture. U.S. Government Accountability Office.

Matthew Fredrikson, Eric Lantz, Somesh Jha, Simon Lin, David Page and Thomas Ristenpart, 2014. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In 23rd USENIX Security Symposium (USENIX Security 14): 17-32.

Andrew Geil, Glen Sagers, Aslihan D. Spaulding and James R. Wolf, 2018. Cyber security on the farm: An assessment of cyber security practices in the United States agricultural industry. International Food and Agribusiness Management Review (3): 317-334.

Maanak Gupta, Mahmoud Abdelsalam, Sajad Khorsandroo and Sudip Mittal, 2020. Security and privacy in smart farming: Challenges and opportunities. IEEE Access (8):34564-34584.

Jamie Hayes, Luca Melis, George Danezis and Emiliano De Cristofaro, 2019. Logan: Membership inference attacks against generative models. Proceedings on Privacy Enhancing Technologies 2019(1): 133-152.

Nicolas Papernot, Patrick McDaniel, Arunesh Sinha and Michael Wellman, 2016. Towards the science of security and privacy in machine learning. Arxiv 1611.03814.

Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik and Ananthram Swami, 2017. Practical black-box attacks against machine learning. Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security: 506-519.

Ahmed Salem, Yang Zhang, Mathias Humbert, Pascal Berrang, Mario Fritz and Michael Backes, 2018. Ml-leaks: Model and data independent membership inference attacks and defense son machine learning models. Network and Distributed Systems Security (NDSS) Symposium 2019.

Reza Shokri, Marco Stronati, Congzheng Song and Vitaly Shmatikov, 2017. Membership inference attacks against machine learning models. 2017 IEEE Symposium on Security and Privacy (SP).

Liwei Song, Reza Shokri and Prateek Mittal, 2019. Membership inference attacks against adversarially robust deep learning models. 2019 IEEE Security and Privacy Workshops (SPW): 50-56.

Michael Sykuta, 2016. Big data in agriculture: Property rights, privacy and competition in ag data services. The International Food and Agribusiness Management Review (19): 57-74.

Florian Tramèr, Fan Zhang, Ari Juels, Michael K. Reiter and Thomas Ristenpart, 2016. Stealing machine learning models via prediction apis. In Proceedings of the 25th USENIX Conference on Security Symposium (SEC'16): 601–618

Jason West, 2018. A prediction model framework for cyber-attacks to precision agriculture technologies. Journal of Agricultural and Food Information 19 (4):307-330.

Marc Window, 2019. Security in precision agriculture: Vulnerabilities and risks of agricultural systems. Luleå University of Technology - Department of Computer Science, Electrical and Space Engineering.

Sjaak Wolfert, Lan Ge, Cor Verdouw and Marc-Jeroen Bogaardt, 2017. Big data in smart farming - a review. Agricultural Systems (153):69-80.

Yuheng Zhang, Ruoxi Jia, Hengzhi Pei, Wenxiao Wang, Bo Li and Dawn Song, 2020. The secret revealer: Generative model-inversion attacks against deep neural networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): 250-258.

Xuezhou Zhang, Xiaojin Zhu and Laurent Lessard, 2020. Online data poisoning attack. Proceedings of the 2nd Conference on Learning for Dynamics and Control, PMLR (120): 201-210.