Case Analysis #4 - Data Mining Analysis at your Company/Org

Husky Trio: Vicky, Shouyang, Billy

Data Mining Application 1: Biosecurity and Livestock Health Management

Biosecurity, in the context of livestock farming, refers to measures and integrated approaches to prevent and minimize the spread of infectious diseases that are detrimental to animals. Monitoring livestock health is essential to ensure the safety of the food supply. One study suggests that *PRRSV, a virus that causes stillborn piglets, costs US farmers an annual $600 million*. (USDA, 2020) It is evident that biosecurity is a top priority that Smithfield must take care of before anything else.

One of the significant issues farmers face with monitoring farm livestock is the sheer volume of animals. When a family pet is sick, owners can often notice the abnormality immediately because pets are showered with attention constantly. However, the same is not valid for family farm animals. It is challenging for one or two farmers to monitor all livestock while giving each animal the same amount of attention. Imagine the following scenario:

Gert is a family farmer who owns a dozen pigs, and she has to feed the pigs daily while taking care of other farm chores, such as collecting eggs, hauling water, and tending the garden. Just like Smithfield and other small family farms, Gert doesn't have the money to hire outsiders. The farm is barely sustainable, and the amount of work is taking a toll on Gert. A study by Agriland suggests that *72% of farmers say they work more than a 60-hour week.* (Browne, 2016) As we can observe here, it is not possible for Gert to spend most of her time monitoring the pigs' health. Imagine if you were Gert and the following picture is what you see when you're interacting with the pigs daily.

There are answers we can't obtain by merely observing the animals briefly. For instance:

1. Are there sick pigs? Which animals are sick?

2. Are the sleeping pigs sick, or are they just taking a nap?

3. How active are those pigs?

4. Which pigs suffer from malnutrition?

It is tough to answer those questions without quantitative data. One day, Gurt observed that some pigs are dead, but it's too late to do anything because the disease has already spread. Gurt ended up losing all of the pigs, and the financial consequence is devastating.

How can farmers prevent the disaster from happening?

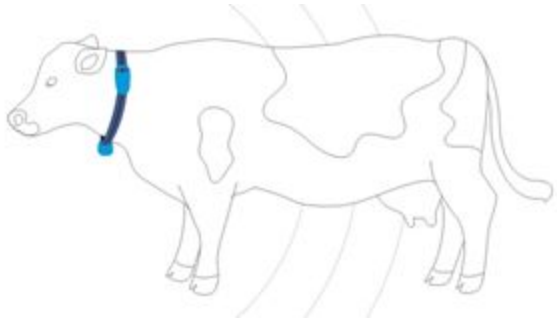
With the help of AI sensors and cameras, data mining is a robust process that farmers can leverage to monitor their livestock's health. There are many warning signs that an animal is sick, and it is integral for farmers to catch those signs immediately. However, Gert and many other farmers' main problem is that they often fail to recognize the warning signs because there are way too many animals for one individual farmer to monitor. A human can't observe an animal 24/7, but a machine could.

Real-time sound monitoring – finding anomalies within large sound data sets

In their paper, "*Recognition of sick pig cough sounds based on convolutional neural networks in field situations*," Yin et al. described the application of data mining in the detection of livestock diseases. There are many valuable insights farmers can gain from analyzing the sounds an animal makes. The paper suggests that *sounds issued by pigs can be analyzed for the detection of diseases. In particular, their coughs can be studied because they indicate their sickness*. (Yin et al., 2020) By setting up microphones in various positions, the AI system can differentiate different sounds that pigs make and warn the farmer once an abnormality is recorded. This is especially helpful when it comes to the early detection of diseases. By setting up a smart camera, the system can locate the sound source and find patient zero in the matter of a second to avoid further disease transmission from happening. The authors conclude that *the results reveal that the proposed algorithm significantly outperforms the other algorithms—cough and overall recognition accuracies reach to 96.8% and 95.4%, respectively, with 96.2% F1-score achieved*. (Yin et al., 2020)

Real-time rumination (chewing) monitoring – finding eating patterns within large data sets

Another significant warning sign of a sick animal is a lack of appetite. When a cow isn't feeling well, it will eat significantly less than usual. It is challenging to notice this eating habit if the farmer doesn't pay close attention to the animals. Afimilk, an innovative farm equipment company, offers a neck collar that has the ability to monitor the health of individual cows.



Placing a smart collar on the cow's neck analyzes the cow's chewing pattern and sends the data back to the farmer's computer. The user will receive an alert when an abnormality occurs, and *the alerts detail the changes and possible causes*. (Afimilk, 2021) Furthermore, each device can be traced back to a unique animal ID. The information about animal health is presented on AFimilk's farm management software, which ensures that the monitoring process is automated with minimal human intervention.

By applying advanced AI technology and data mining techniques, farmers can monitor all livestock's health at ease while reducing workload. Moreover, detecting and separating sick animals effectively and promptly ensures biosecurity by preventing the disease from spreading at an early stage.

Challenges Concerning Data Mining and Biosecurity

Data mining in livestock health management is a very recent research topic, far from being matured. It is fairly reasonable to assume that most family farmers don't have the educational background or technological vision to leverage a comprehensive BI and data mining system. Because of the technology's untraditional nature, many farmers don't know it well enough to justify the extra financial cost. Even if a smart collar costs less than $100, the price will add up as the farmer implements it on all farm animals. Many traditional farmers will avoid data mining entirely because they don't trust the technology. Your data mining system will not be successful if you don't even have a place to "mine your data" in the first place.

Furthermore, due to the technology's novelty nature, the sensors and devices still need more research. Yin et al. note that misclassification is an area that can be further improved when it comes to identifying the sounds animals make. The cough can be further classified "into dry/wet coughs, single/continuous coughs, or strong/weak coughs which have a certain significance in clinical diagnosis" to be more accurate. On top of that, a lack of scientific research is also a problem for other smart sensors. Animals vary in size and features. The ability to monitor an animal effectively requires years of clinical trials and concrete quantitative data to back it up.

It is relatively easy to find patterns and abnormalities within large data sets. However, data mining findings are meaningless if the data sets are inaccurate, incomplete, or inconsistent. Many steps could tamper with the integrity of the data. Different algorithms could potentially generate false-positive or false-negative results. Technology and a lack of tech vision remain the main challenges when it comes to the success of data mining regarding biosecurity.

Data Mining Application 2: Precisely Interpret Data from the Internet of Things (IoT)

The Smithfield Meat Farm raises a great variety of livestock that is grass-fed and free-range, including sheep, cattle, pigs, etc. Just like all the current smart farms, it is essential for Smithfield to implement the Internet of Things that can keep track of all kinds of activities that are related to the livestock and its living and grazing environment. With a large amount of data constantly extracted from the IoT system such as collars and ultrasound sensors, Smithfield must engage in a proper data mining process to precisely interpret the data and unveil unknown patterns so that it can make powerful data-driven decisions.

Understanding Animal's Grazing Behavior and Providing Valuable Insights

As we mentioned previously, a majority of livestock in Smithfield is free-range, indicating that it usually requires a great effort for farmers to keep track of all the livestock and make sure every single animal is within a desired geographical boundary. In addition, it has been observed that livestock such as cows and buffaloes are sluggish as compared to goats and sheep: they tend to have lower physical excitement after going through the grazing period and thus prefer to have rest or lay down for a long time (Ilyas and Ahmed, 2020). Therefore, with sensors that can track the GPS location of every livestock, farmers can retrieve data of the geographical motion and then use data mining algorithms to understand grazing behaviors of different species of animals.

A direct benefit from this process is that farmers can have their workload significantly reduced since by understanding the physical actions of different animals they can effectively manage the herd using an end-user device and publish efficient grazing schedules for different livestock. In addition, by utilizing the IoT system and understanding the data mining insights, Smithfield can effectively reduce waste associated with any loss of livestock and lower the cost of unnecessary utilization of resources.

Deciding on Proper Maintenance of Grass

Since many of Smithfield's livestock is grass-fed, the quality of grass directly affects the quality of meats and other products. Effective monitoring of the grass quality can be done with a combination of a video camera that records the grass field and sensors on livestock that track animal's eating behavior. In research by Nóbrega et al. (2018), sheep were released onto a plain field and their activity recorded on video for about 3 hours, while collars continuously retrieved time-stamped raw sensor data and sent it into the network in order to be manually classified. This was then processed by a machine learning algorithm to predict the extent of influence on grass for a specific number of sheep. Here, data mining methods are critical since all data received from different devices must be carefully processed to find patterns that are valuable for farmers.

With this system, Smithfield can be sustainable in terms of grazing and the maintenance of the grass. Moreover, by keeping its grass quality at high standards, Smithfield can ensure its product quality and plan for the pricing and the budget more efficiently and systematically.

Challenges with Integrating IoT and Data Mining

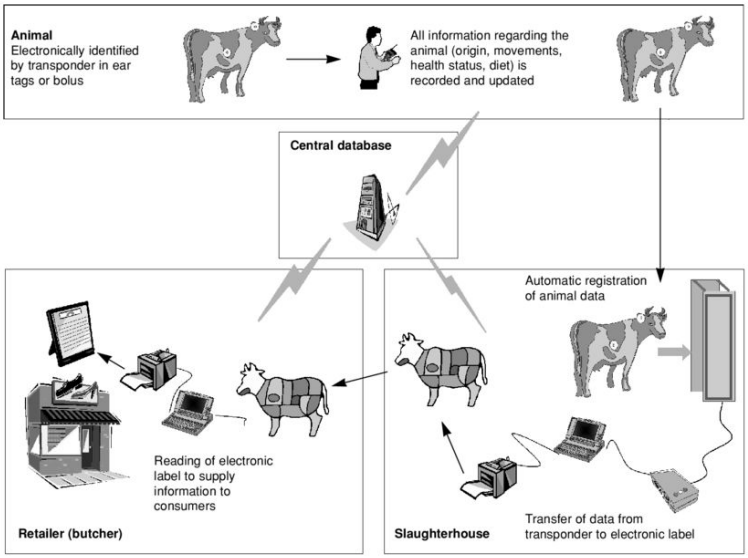
A significant issue in data mining with the Internet of Things is the efficiency of storage and extraction of data. While having multiple sources of data with the implementation of IoT, the amount of data can be TB (terabytes), even PB (petabytes) and ZB (zettabyte) (Chen et al., 2015). Therefore, Smithfield needs to explore fast and effective mechanisms to successfully manage the chaos of data.

Another problem associated with the IoT is the variety of data types. As we showed, IoT in a farm can involve all kinds of devices such as a camera, collar, GPS sensor, etc. As a result, Smithfield must integrate raw data in different formats from different systems in order to go through the data mining process.

In addition, successful implementation of data mining and the IoT system requires well-built and mature infrastructure such as high-speed and stable internet access. For a traditional and home-own farm like Smithfield, this is likely to be costly and thus requires careful budgeting in order to efficiently utilize the whole system to make revenue.

Data Mining Application 3: Traceability of livestock products

Back in early 1000 BC, farmers developed a simple traceability system by using iron and ear cuts to mark livestock to prevent theft. And in 2021, Smithfiled shall use the power of data mining to establish its own traceability system to record the growth environment of livestock, ensure that the food product they sell is organic and safe, and be able to track and recall products when there is an issue. Data mining is the process of finding useful information from a large number of data that can be useful and applicable to real life situations. Data mining is a very powerful tool that combines machine learning, mathematical modeling, visualization and statistical techniques to help find and predict data for the users. Techniques such as, association, classification, and clustering, help users to build models that find the relationships between variables. By using a powerful tool like data mining, owners of Smithfiled can build a traceability system for their products. For instance, a mongolia scientist *used RFIDs to trace, track and monitor cattle individuals and built up an end-to-end tracking system embedded with a pre-warning function for cattle breeding and chilled-fresh beef production based on networking and database technologies.(Zhao, Li, Jin, & Pan, 2020)*



Benefits:

Tracking the process of agricultural products from farmland to customers helps prevent diseases, reduce pollution and increase revenue. *Australia has built a permanent animal identification system, the National Livestock Identification System (NLIS), which is able to track animals from their birthplace to the slaughter house. Ear tags, ruminal bolus, and cattle passometer, identified with NLIS certified, need to be instantaneity recorded into the NLIS database by RFID readers which are installed in cattle farms, sale yards and slaughterhouses* (Tonsor & Schroeder, 2004). As the global supply chain becomes longer, the importance of tracking and monitoring agricultural products is also growing. Data mining can improve the quality of operations in farmland and provide a traceable system for all the customers who purchase their products. By utilizing tools such as sensors, scanners and other analytical technology during daily operation, Smithfield can monitor the daily activities of livestock and other crops, collect association data and store the data in a centralized database. During transportation, the temperature and humidity are monitored in real time through sensors with GPS functions. When the requirements are not met, an early warning will be issued for correction. Therefore, the sales lead can take immediate and efficient response measures when there is a problem or need to recall food, and even after the product is sold. The good part about data mining is that it provides tools by which big data can be analyzed automatically.

Challenges

The major challenge for Smithfiled to implement the traceability system will be the high cost. First of all, there are lots of different veterinary medications/drugs in the market that can greatly improve the food supply and reduce the product and operational cost. Smithfield can easily take the risk to use the drugs to improve the health and fertility of their livestock to gain more profit.

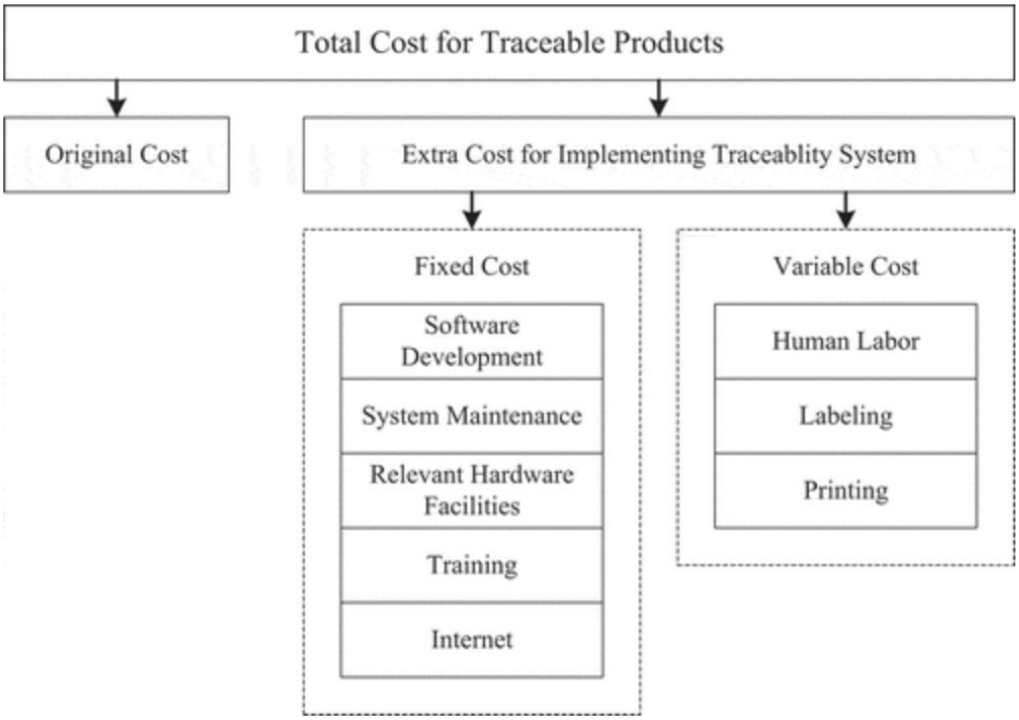
However, to ensure the food source safety and provide the best service for customers, Smithfield chose a different approach. See below for the detailed cost of livestock identification methods.

Table 2. The price, loss and readable rate in livestock identification.

Identification	Livestock	Location	Price	Period	Loss rate	Readable rate	Data cited
2D barcode tag	Swine	Ear	$0.01	180d	5%	86%	(Lu et al., 2004)
Rumen bolus	Cow	Rumen	$5.0	300d	0.3%	99.7%	(Fallon, 2001)
RFID tag	Goat	Armpit injected	$5.0	180d	1.9%	98%	(Caja et al., 1999)
Electrical tag	Beef	Ear	$8.0	300d	1%	99%	(Basarab, Milligan, & Thorlakson, 1997)
RFID tag	Cow	Ear	$2.0	180d	3%	93%	(Bai, 2010)
Rumen bolus	lamb	Ear	$4.0	77d	3.2%	100%	(Garín et al., 2003)
Retina scan	Cattle	Eye	$100	perpetual	–	100%	(McMahon, 2000)
DNA fingerprint	Cattle	Blood	$100	perpetual	–	100%	(Negrini et al., 2008)

(Bai et al., 2017)

We can see that it can cost up to $100 per cattle for retina scan method, and tag costs vary from $2 to $8 depending on the identification practices. *The total cost for implementing NAIS in the cattle sector as described in the study is $175.9 million annually (at a 90 percent participation level).(*Gómez-Lobo, 2012) There are chinese companies that spend almost 2 million dollar on developing traceability software, and as we always say, we recommend Smithfiled to purchase software instead of building it, and to purchase such software.Thus, the high cost of implementing such system is definitely the core challenge for Smithfield. However, other issues such as maintenance and reliability of the system also need to be noted. For example, the system needs quarterly or yearly inspection, update and upgrade to be at its optimal state, and the cost of hiring technicians and maintenance personnel can't be overlooked. More importantly, it will cost Smithfiled a fortune if the system is not reliable, such as false alarm, sending wrong signal to recall the product that has no issue. Then Smithfiled suffers from shipping to replacement cost and the worst reputation stigma that may influence the future of the company.



(Chen, Tian, & Xu, 2018)

References

Animal Health | National Institute of Food and Agriculture. (2019). Retrieved from Usda.gov website: https://nifa.usda.gov/topic/animal-health

Browne, E. (2016, July 29). Here's how many hours farmers REALLY work in a week. Retrieved from Agriland.ie website: https://www.agriland.ie/farming-news/heres-how-many-hours-farmers-really-work-in-a-week/

Y, Y., & T, D. (2020). Recognition of sick pig cough sounds based on convolutional neural network in field situations. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2020.11.001

Tonsor, G. T., & Schroeder, T. C. (2004). Australia's livestock identification systems: implications for United States programs. Australia's Livestock Identification Systems: Implications for United States Programs.

Bai, H., Zhou, G., Hu, Y., Sun, A., Xu, X., Liu, X., & Lu, C. (2017). Traceability technologies for farm animals and their products in China. *Food Control*, *79*(0956-7135), 35–43. https://doi.org/10.1016/j.foodcont.2017.02.040

Chen, H., Tian, Z., & Xu, F. (2018). What are the cost changes for produce implementing traceability systems in China? Evidence from enterprise A. *Applied Economics*, *51*(7), 687–697. https://doi.org/10.1080/00036846.2018.1510470

Ilyas, Q., Ahmad, M. "Smart Farming: An Enhanced Pursuit of Sustainable Remote Livestock Tracking and Geofencing Using IoT and GPRS", Wireless Communications and Mobile Computing, vol. 2020, Article ID 6660733, 12 pages, 2020. https://doi.org/10.1155/2020/6660733

Nóbrega, L., Tavares, A., Cardoso, A., Gonçalves, P., "Animal monitoring based on IoT technologies", 2018 IoT Vertical and Topical Summit on Agriculture - Tuscany (IOT Tuscany), 2018.

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. International Journal of Distributed Sensor Networks. https://doi.org/10.1155/2015/431047