Ensuring the Integrity of the European food chain

Early Warning System for Food Fraud Detection: Machine Learning Applied to Food Big Data

Vahid Mojtahed (vahid.mojtahed@fera.co.uk) – Fera Science Ltd, National Agri-food Innovation Campus, Sand Hutton, York YO41 1LZ

INTRODUCTION

On average, nearly 30 product recalls occur every week in Europe with the costs between 2-15% of a company's yearly revenues. The current intervention systems are not capable of uncovering the near infinite number of contaminants or ingredients used for food fraud, and the nature and instances of food fraud are constantly changing. We have applied advanced analytics such as Machine learning as well as statistical analysis to food Big Data in order to develop an Early Warning System for monitoring food supply chains and detecting anomalies as a sign of food fraud activities. This approach is truly agnostic to prior information thus effective in detecting emerging risks.

a) Early Warning Systems (EWS)

There is a need for cost-effective proactive methods such as EWS in contrast to reactive methods such as detection and interventions based on analytical tests, which are costly and time consuming for mitigating food fraud risks. The following graph shows where EWS sits in the time and cost line of food fraud management.

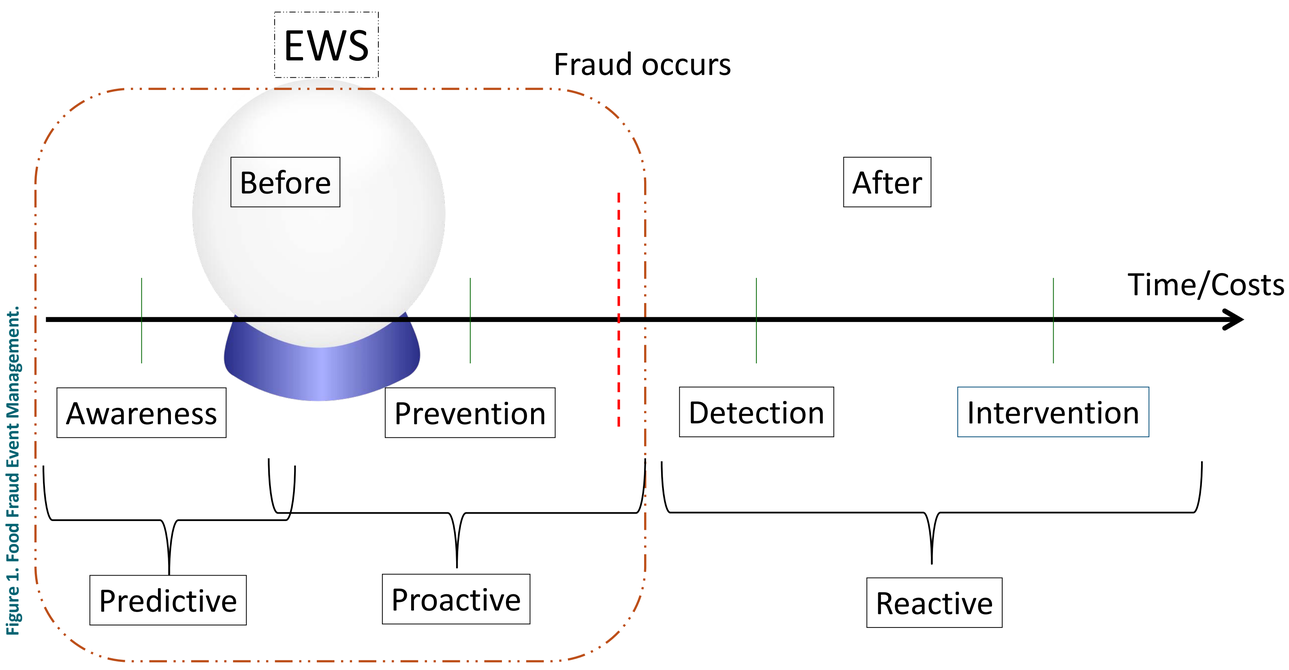

Figure 1. Food Fraud Event Management.

b) Advanced Analytics and Big Data

Detecting fraud, in real-time, requires monitoring and detecting subtle changes in few or many relevant triggers of food fraud. These triggers are context-based and are chosen given the commodities characteristics. A taxonomy of the comprehensive set of these triggers are provided in Figure 3. Advanced analytics such as combination of Machine learning and statistical analysis are jointly used to identify anomalies in the combined dataset (Food Big Data). The identified anomalies are carefully examined against reference distribution to establish their genuineness as extreme events such as food fraud incidences. This approach is agnostic to prior (i.e. historical) information regarding the nature of fraud, types of commodities, and country of origin.

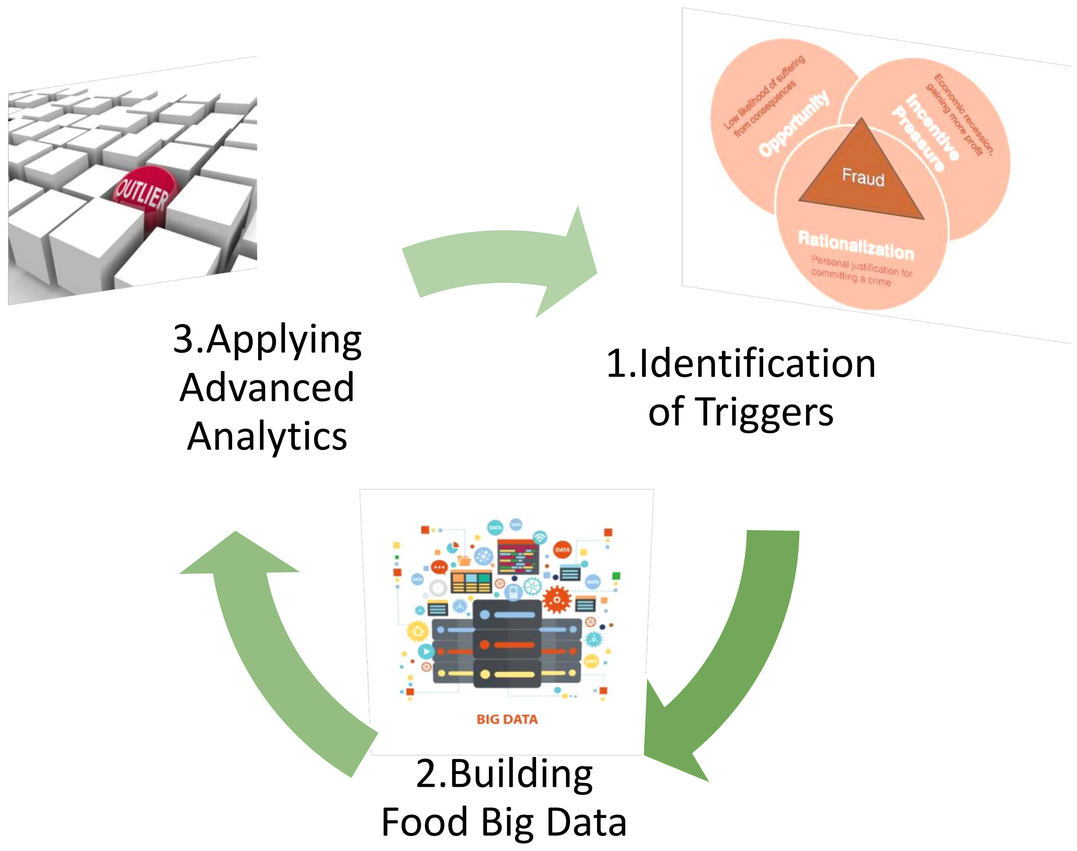

Figure 2. EWS components and flowchart..

1 **Socio-economic characteristics**

- Relative prices of ingredients/inputs
- Price of final product
- Consumers' behaviour (preferences, attitude, consumption pattern, etc.)
- Demand characteristics
- Excess demand/short supply with lags (previous year demand or current year supply)
- Economic downturn (Dennis & Kelly, 2013)
- Tariffs, taxes, and custom duties - i.e. country of origin mislabelling (China) for avoiding anti dumping duties

2 **Regulations and Standards**

- Penalties
- Change in regulations and standards – i.e. Ochratoxin A from 30 µg/kg to 20 µg/kg for spices

3 **Industries' or sectors' characteristics**

- Profit margin (Gross value added as a proxy)
  - Low profit margin driving fraud in high volume products
  - High profit margin driving fraud in high premium products
- Supply chains length and complexity
- Production capacities

4 **Commodities' characteristics**

- Durability
- Substitutability - i.e. pomegranate juice with apple/grape juice

5 **Climate/Pests/Livestock Health/Labs**

- Climate variation- i.e. addition of Sudan dye when paprika has mould
- Vulnerability to pests or visits to vets
- Lab samples and results

Figure 3. Taxonomy of Food Fraud Big Data

c) Results

The EWS concept and methodology have been successfully tested and validated retrospectively in case studies such as well-known Horsemeat scandal. In the Horsemeat case study, using only two types of data i.e. quantity and prices, the EWS was able to detect two anomalies in international trade of horse meat that was not compatible with normal pattern of data. The EWS in this case could have raised red flags of fraudulent activities up to 6 months before the outbreak of the scandal in January 2013!

d) Future Research: Fraud Diffusion in Supply Chain

The food supply chain is complex and it is constantly evolving. The Horsemeat scandal in 2013 was one example of the global interwoven network of commodity trades. To forecast the emerging risks, we need to map the underlying structure of food chain across countries and commodities to enhance our understanding of the risk propagation in food supply network and identify systematic fraudulent activities.

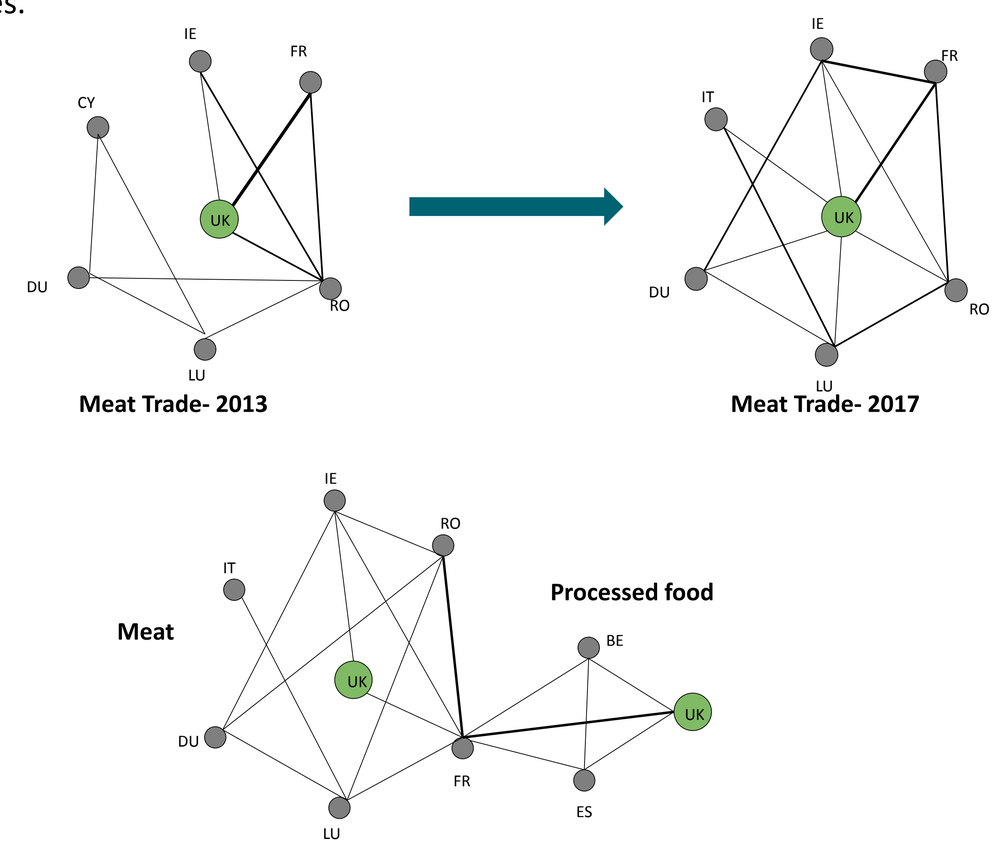

Figure 4. Diffusion of Fraud Risk in Global Food Supply Chain.

The project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 613688.