

The use of Big Data in food safety

大数据在食品安全中的应用

Hans Marvin, Yamine Bouzembrak, Martijn Staats and Esmée Janssen (RIKILT Wageningen UR)

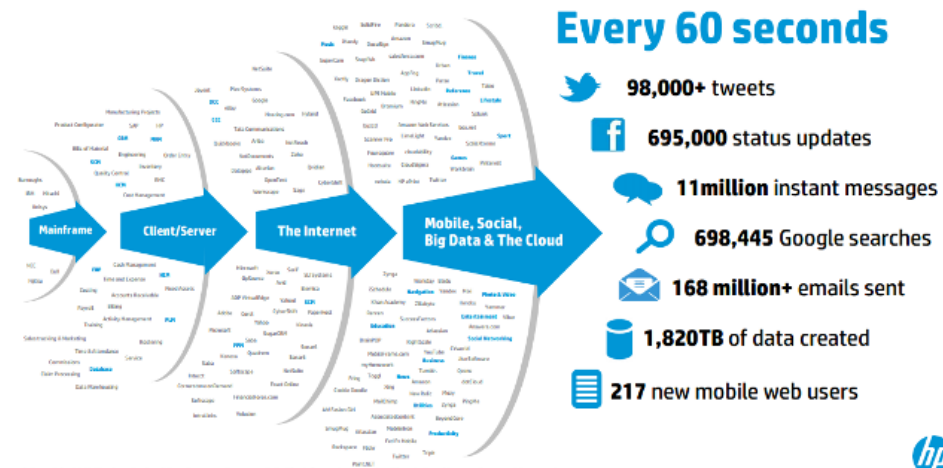


Outline

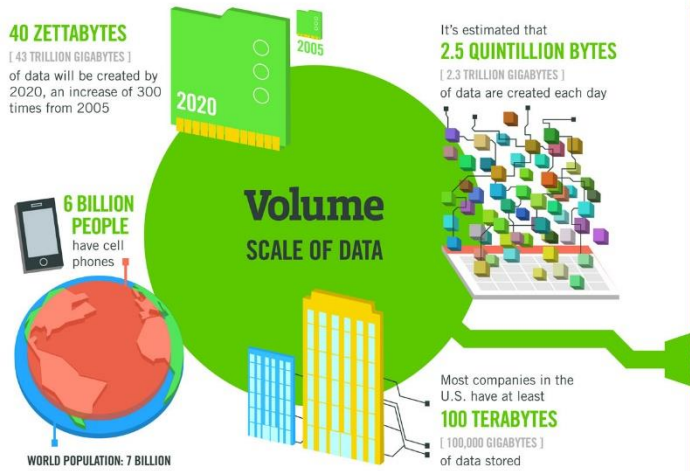
- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Introduction

- Lots of data is being collected and stored, e.g.:
 - Web data, e-commerce
 - Bank/Credit Card transactions
 - E-mail
 - Data recorded by sensors, scanners, other mobile devices.
 - Chemical monitoring data
 - Omics data



Introduction



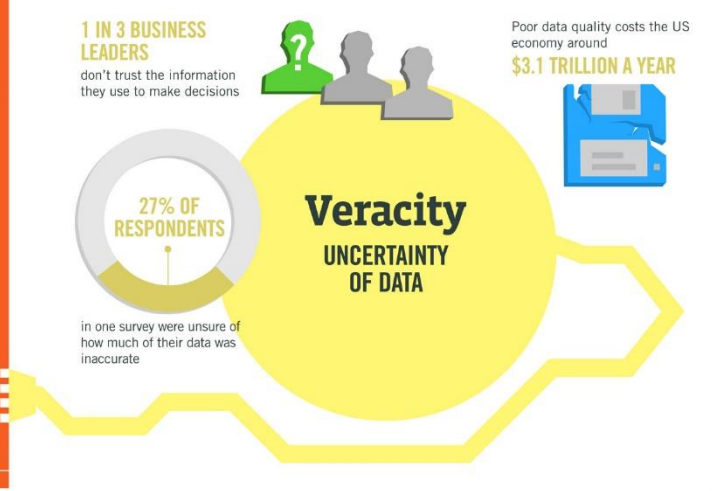
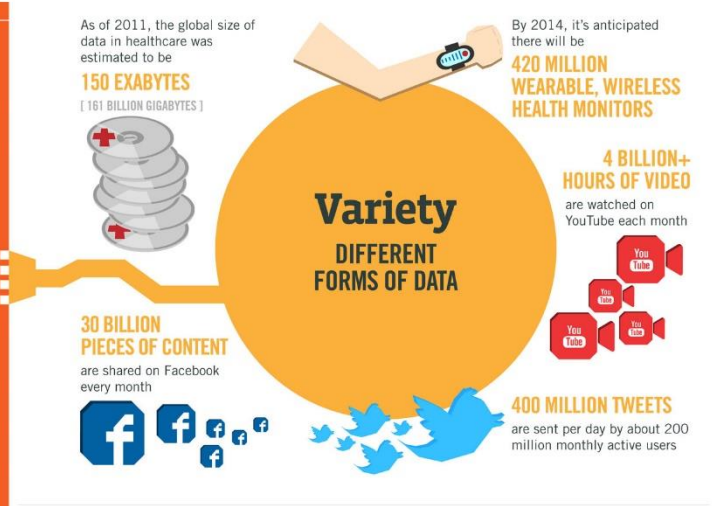
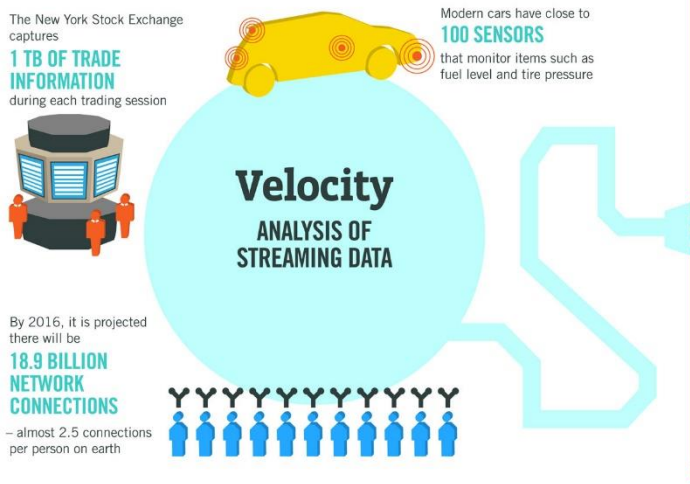
The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015, **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS



Introduction

Definition of European Commission:

The term "big data" refers to large amounts of different types of data produced with high velocity from a high number of various types of sources. Handling today's highly variable and real-time datasets requires new tools and methods, such as powerful processors, software and algorithms.⁷

⁷ Going beyond traditional "data mining" tools designed to handle mainly low-variety, small scale and static datasets, often manually.

Creating value
out of Big Data

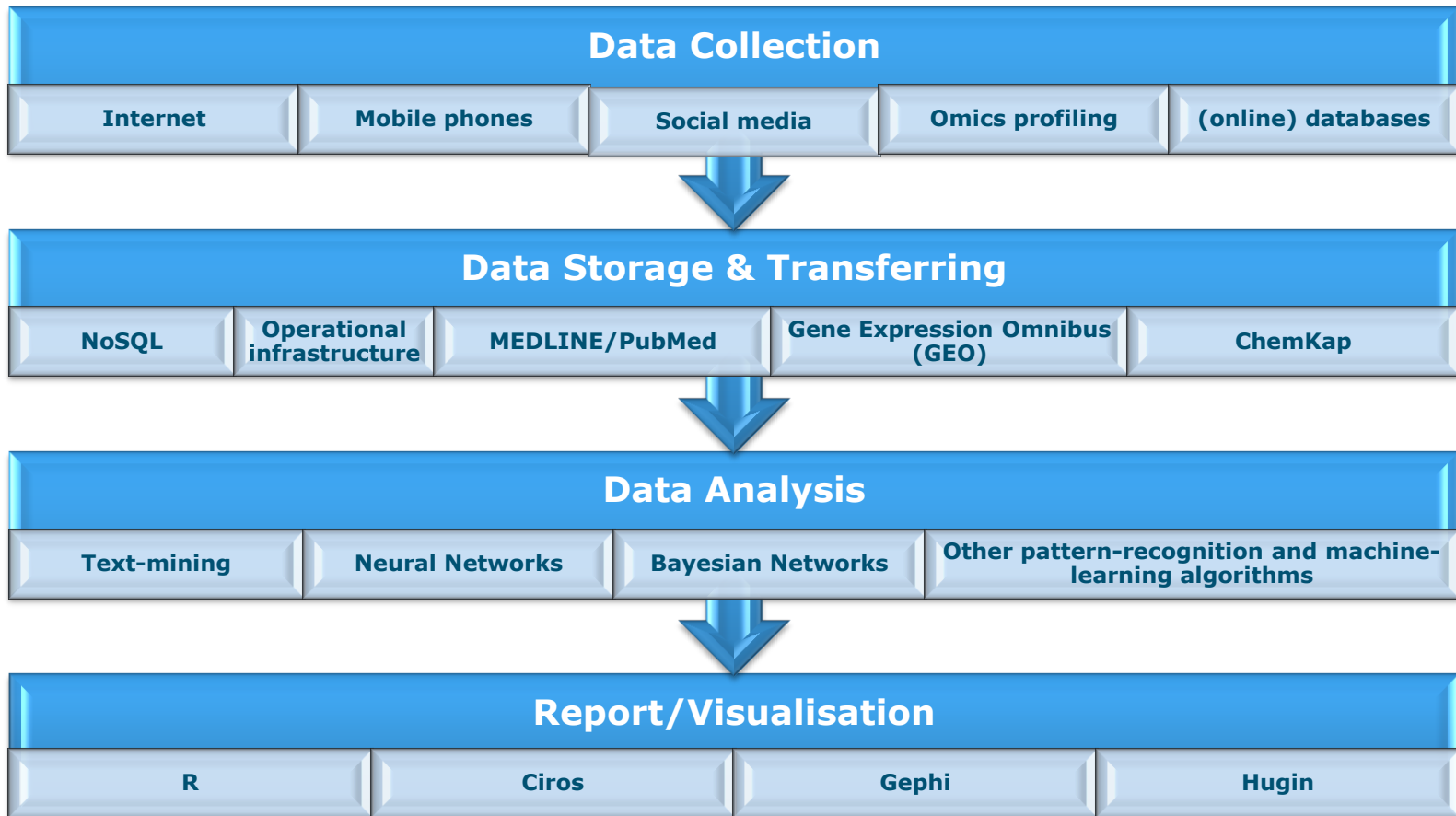


Introduction

- The EC has developed a strategy on Big Data and supports a data-driven economy.
 - Big Data as the centre of the future knowledge economy.
 - Supporting open access of data, global connections and e-infrastructure services (www.openaire.eu)
- The Dutch Government stimulates public-private projects to explore Big Data.
 - Big Data Value Center (www.bdvc.nl)
 - Netherlands eScience center www.esciencecenter.nl

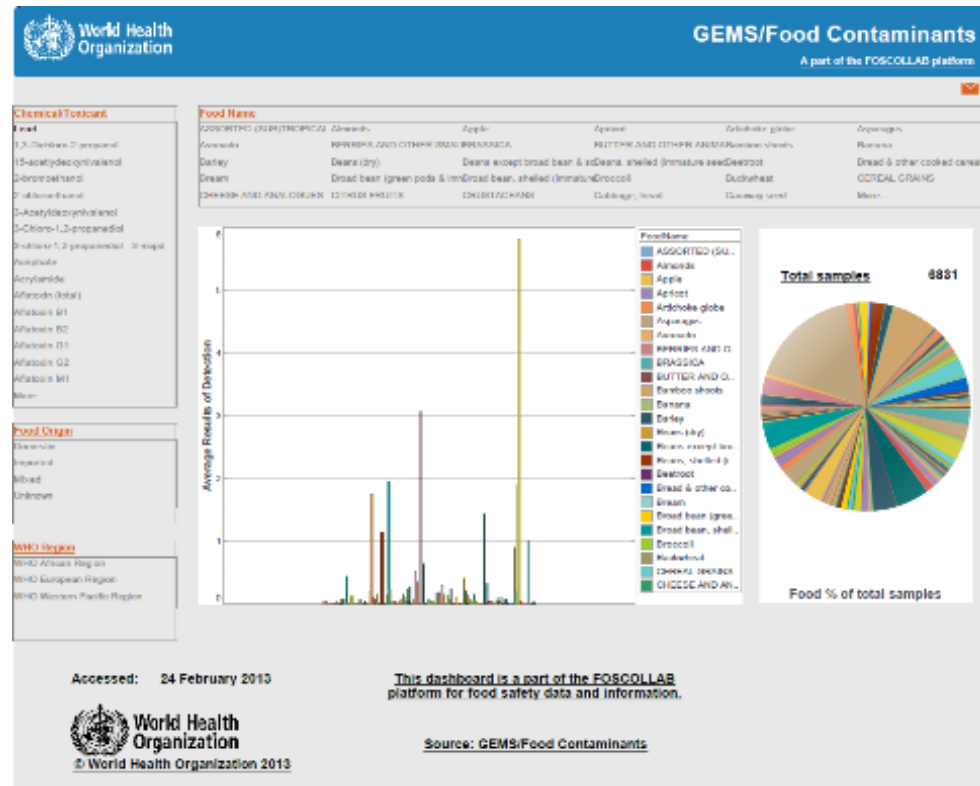
Introduction

Typical Big Data workflow



Introduction; example Big Data in food safety

- FOSCOLLAB: Global platform for food safety data and information.
- Support risk managers and decision-makers
- Integrate data and information from animal/agricultural, food, and human health
- Promote better data generation



Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Use of media in food safety alerts

The Europe Media Monitor (EMM) provides advanced analysis systems for monitoring of both traditional and social media.

- EMM presents the latest news and classifies it according to subject.
- It is updated every 10 minutes, 24 hours per day.
- It gathers reports from news portals world-wide in 60 languages,
- EMM classifies the articles, analyses the news texts, aggregates the information and issues alerts.
- EMM applies text mining techniques to screen different types of media on the world wide web: websites, databases, blogs, ..etc.
- EMM contain 3 portals: NewsBrief, NewsExplorer and MedISys

Use of media in food safety alerts

MedISys collects food safety news and RIKILT adapted it also to detect food fraud

The screenshot displays the MedISys website interface. At the top left is the MedISys logo, and at the top right is a search bar with the text "advanced search". A left sidebar contains a navigation menu with categories like "Top Stories", "Event Extraction", "Recent Disease Incidents", "Alert Statistics", "Communicable Diseases", "Symptoms", "Bioterrorism", "Nuclear", "Chemical", "ECDC", "EFSA", "Other", "Continents", "Official Sources", and "Sources List". The main content area is titled "Latest News About - FoodIntegrity" and features a pagination control showing page 1 of 10. Below this, there are three news articles categorized by their publication time: "Articles published more than 12 hours ago", "Articles published more than 1 day ago", and "Articles published more than 3 days ago". The first article is titled "Beware Christmas food fraud, warns expert" and is dated Monday, November 3, 2014. The second article is "QMS boss emphasises importance of Scotch meat quality assurance" dated Monday, November 3, 2014. The third article is "Food industry must consider intentional contamination - NSF" dated Friday, October 31, 2014. On the right side, there is a "Tools" section with options for "Tuesday, November 4, 2014 1:10:00 PM CET", "RSS | KML | MAP", "subscribe | manage", "current page | all pages", and "info | definition". Below the tools is a "Languages" section with a grid of language codes (af, am, ar, az, be, bg, bs, ca, cs, da, de, el, en, es, et, fa, fi, fr, ga, ha, he, hi, hr, hu, hy, id, is, it, ja, ka, km, ko, ku, lb, lo, lt, lv, mk, mt, my, nl, no, pap, pl, ps, pt, ro, ru, rw, se, sk, sl, so, sq, sr, sv, sw, ta, tg, th, tr, uk, ur, vi, zh) and an "all" button. At the bottom right, it says "Interface: en - English".

MedISYS

Search advanced search

Latest News About - FoodIntegrity

10 < 1 2 > +10

Articles published more than 12 hours ago

Beware Christmas food fraud, warns expert

foodmanufacture Monday, November 3, 2014 7:57:00 PM CET | info [other]

Manufacturers should be particularly vigilant over the festive season, warned food safety and brand protection firm Qadex. The firm's operations manager Tracey Cranney said: "Having to compete with successful low-price retailers is tough, especially heading into the festive season when food budgets will be cut to allow spending in other areas...."

Articles published more than 1 day ago

QMS boss emphasises importance of Scotch meat quality assurance

meatinfo Monday, November 3, 2014 11:36:00 AM CET [other]

Speaking to around 150 farmers at McIntosh Donald's Producer Club Open Evening, Morton emphasised the value to Scottish livestock producers of having strong brands and robust quality assurance schemes. He said: "The Scottish red meat industry is committed to delivering a top-quality product, but one....."

Articles published more than 3 days ago

Food industry must consider intentional contamination - NSF

foodqualitynews Friday, October 31, 2014 5:47:00 PM CET | info [other]

NSF International is an independent organization that writes standards, and tests and certifies products for the water, food, health sciences and consumer goods industries. Serban Teodoresco, global managing director, NSF's consulting and technical group within the food division, said it was important to have a proactive program in place....

Tools

Tuesday, November 4, 2014 1:10:00 PM CET

RSS | KML | MAP

subscribe | manage

current page | all pages

info | definition

Languages

Select your languages

af	am	ar	az	be	bg
bs	ca	cs	da	de	el
en	es	et	fa	fi	fr
ga	ha	he	hi	hr	hu
hy	id	is	it	ja	ka
km	ko	ku	lb	lo	lt
lv	mk	mt	my	nl	no
pap	pl	ps	pt	ro	ru
rw	se	sk	sl	so	sq
sr	sv	sw	ta	tg	th
tr	uk	ur	vi	zh	
all					

Interface: en - English

RIKILT
WAGENINGEN UR

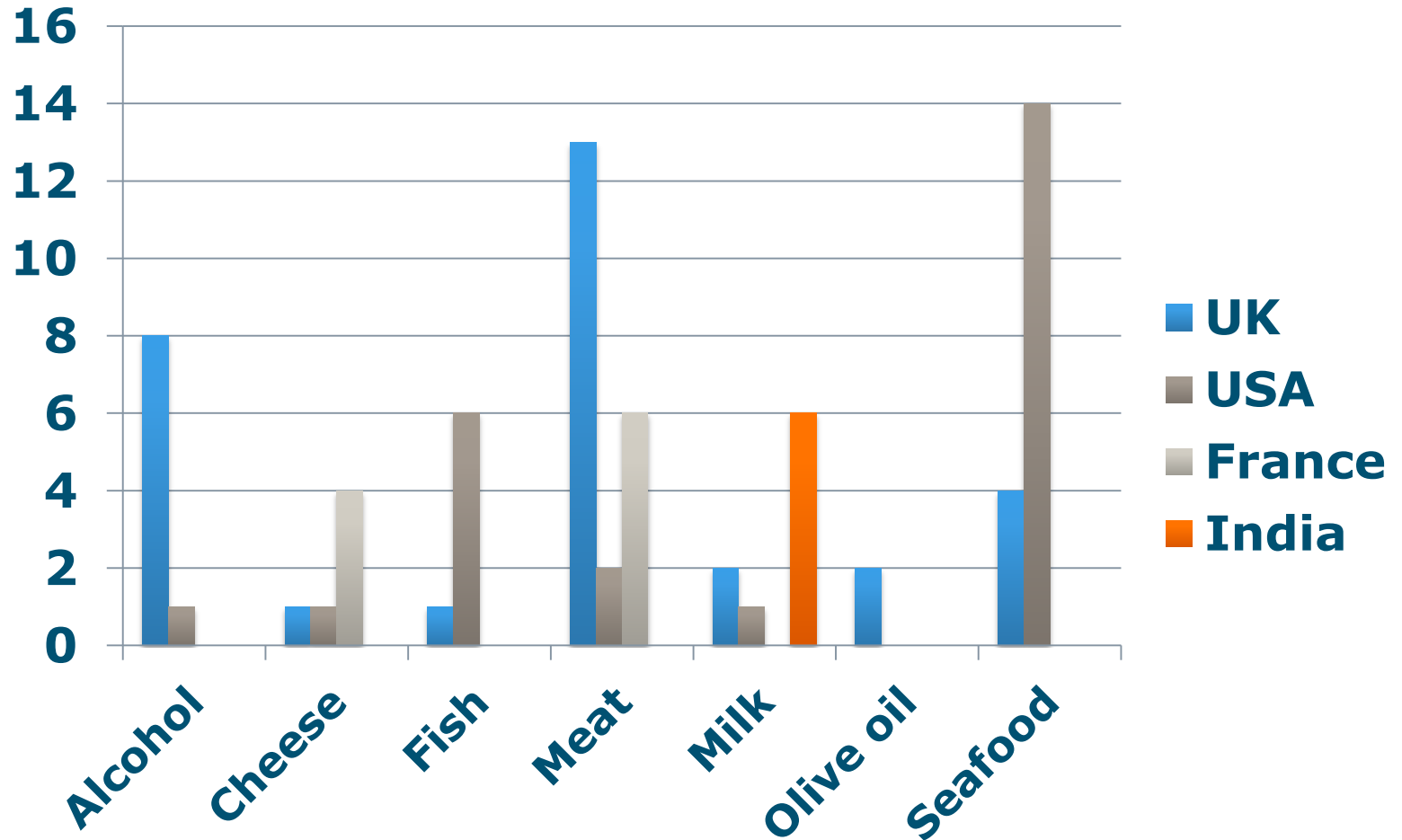
Use of media in food safety alerts

Food fraud reports in MedISys (period September 2014 to June 2015; N > 600)



Use of media in food safety alerts

Food fraud reports in MedISys (period September 2014 to February 2015)



Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Bayesian Networks (BNs) modelling in Big Data research



BN requires less parameters than the conventional naïve method (Dogan, 2012).

BN can combine expert and domain knowledge that allows flexible inference even with partial and limited information (Lauritzen, 1995).

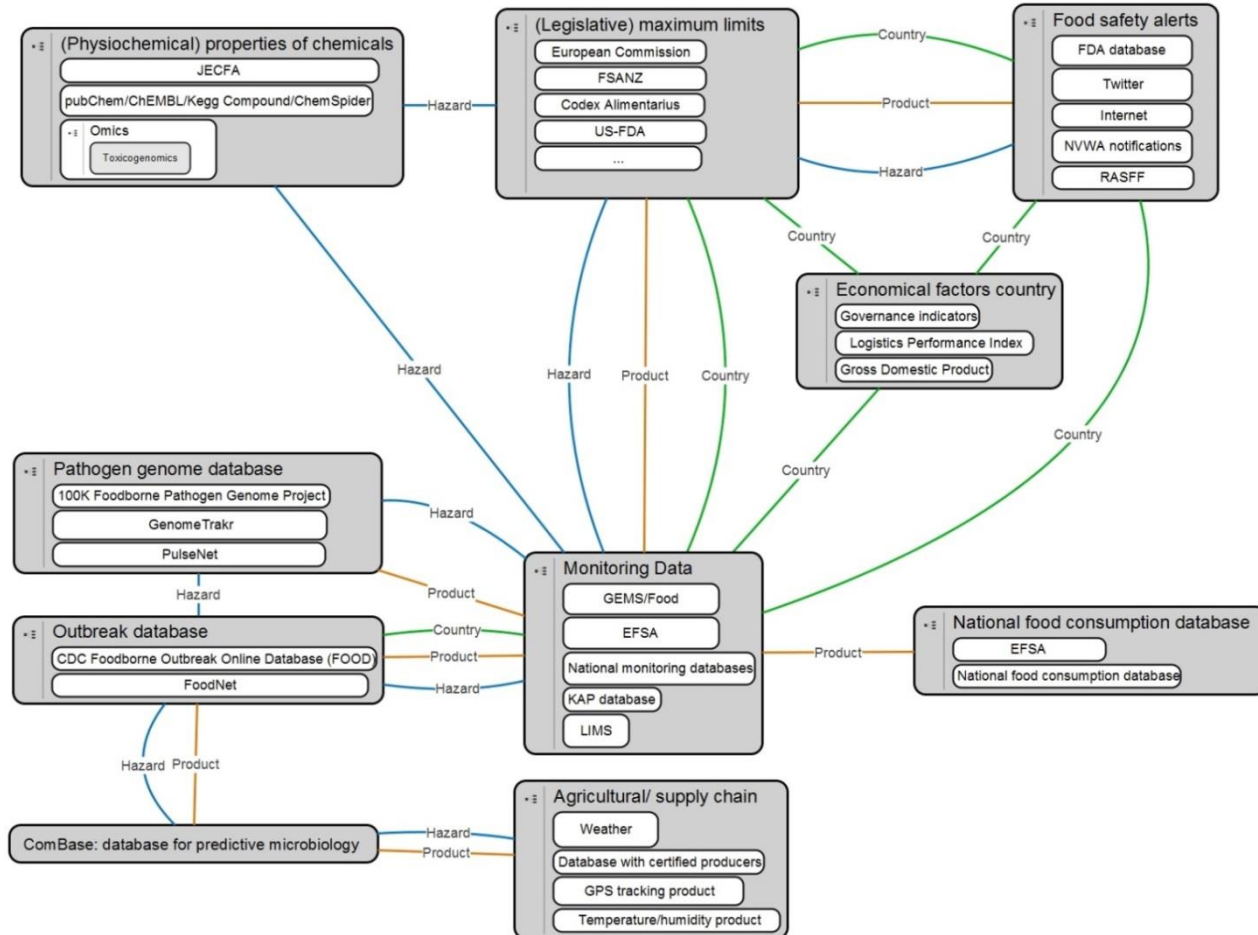
BN are very powerful for making inferences and drawing conclusions based on available information (Jensen, 1996).

BN approach offers a high potential for representing ambiguous knowledge and for performing reasoning under uncertainty

BN takes into account a number of quantitative and qualitative criteria under uncertainty (Dogan and Aydin, 2011)

Bayesian Networks (BNs) modelling in Big Data research

BN is able to combine data from different nature and location



Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Application of BNs to predict food fraud

- (Bouzembrak and *Marvin*, 2015) used BNs prediction model to aid the border inspector to detect the type of food fraud.

ARTICLE IN PRESS

JFCO4660_proof ■ 29 September 2015 ■ 1/8

Food Control xxx (2015) 1–8



ELSEVIER

Contents lists available at [ScienceDirect](#)

Food Control

journal homepage: www.elsevier.com/locate/foodcont



Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling

Y. Bouzembrak, H.J.P. Marvin*

RIKILT Wageningen UR, Akkermaalsbos 2, 6708 WB Wageningen, The Netherlands

Application of BNs to predict food fraud

Prediction of food fraud type in RASFF

RASFF | Consumers Portal | Support | Help | Disclaimer | Log in

European Commission

RASFF Portal

European Commission > RASFF Portal

Notifications list | New search

Search Page

[Get results](#) [Clear form](#)

Notification

Reference

Subject or and

Notified by

Date

Week current week [10]
 previous week [9]
 week of year

Notified between and (dd/mm/yyyy)

Type

Type

Classification withdrawn

Basis

Product

Category

Flagged as

Country

Action taken

Hazard

Category

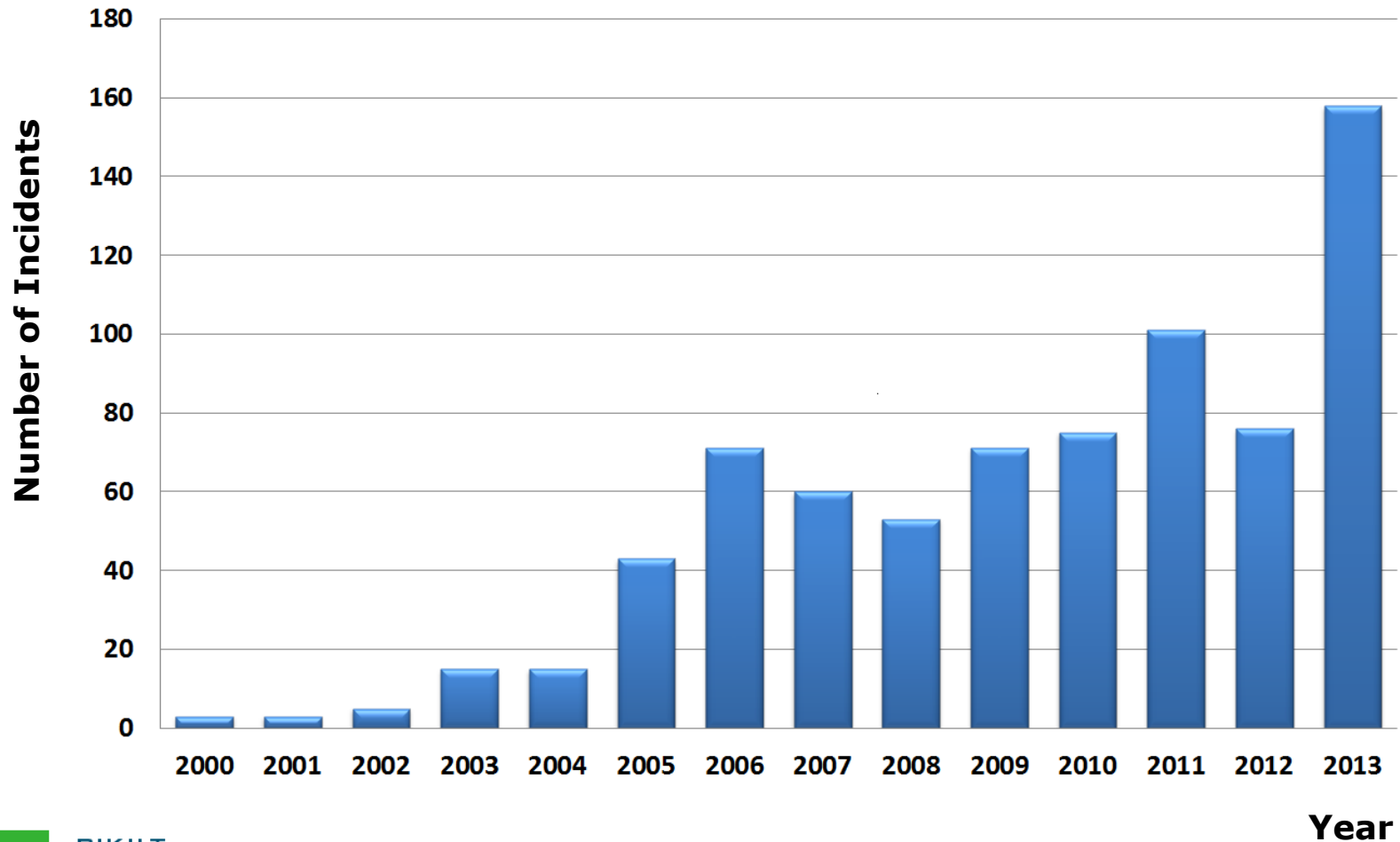
Keywords

Keywords [Open URL](#)

[Get results](#) [Clear form](#) [Load criteria](#) [Save criteria](#)

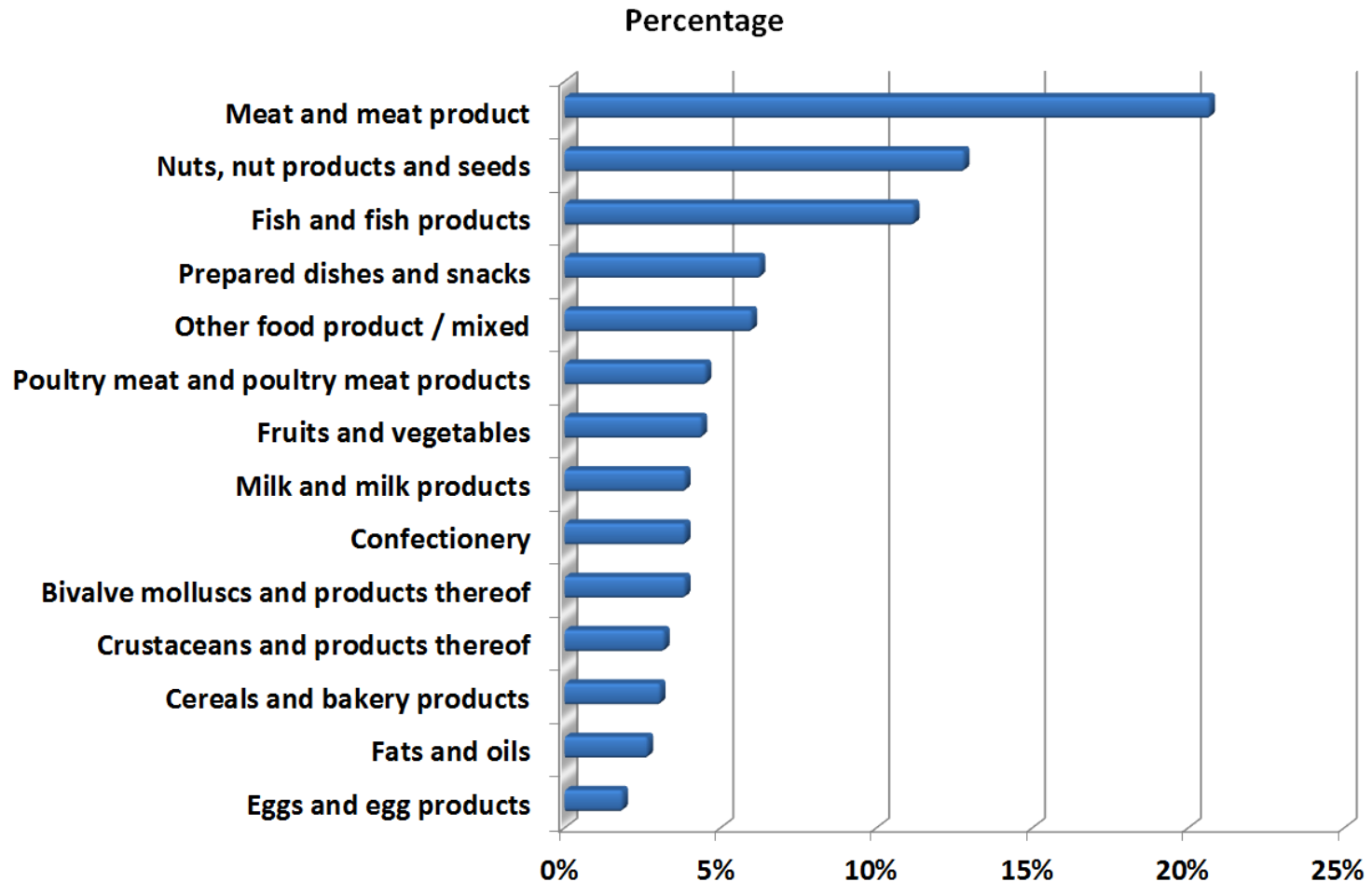
Application of BNs to predict food fraud

Food fraud notifications in RASFF (2000 to 2013).



Application of BNs to predict food fraud

Products related to food fraud in RASFF



Application of BNs to predict food fraud

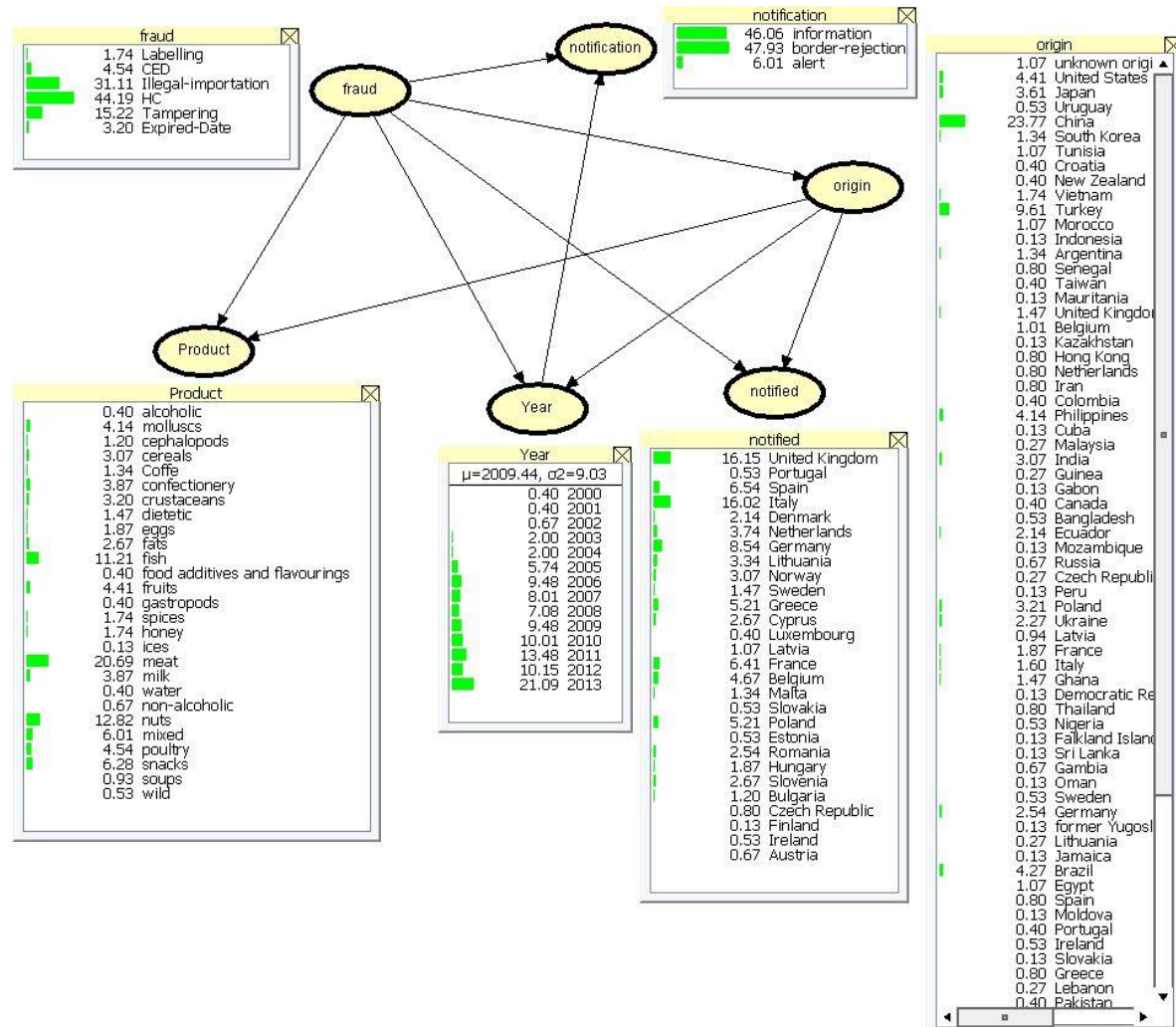
Food fraud types in RASFF

Fraud type	Description	Count of notification type	Percentage
HC	Improper or fraudulent or missing or absent of health certificate	370	44%
Illegal importation	Illegal or unauthorized import or trade or transit	256	31%
Tampering	Adulteration or fraud or tampering	126	15%
CED	Improper or expired or fraudulent or missing of common entry document or import declaration	45	5%
Expired Date	Expired Date	24	3%
Mislabelling	Mislabelling	13	2%
Grand Total		834	100%

Application of BNs to predict food fraud

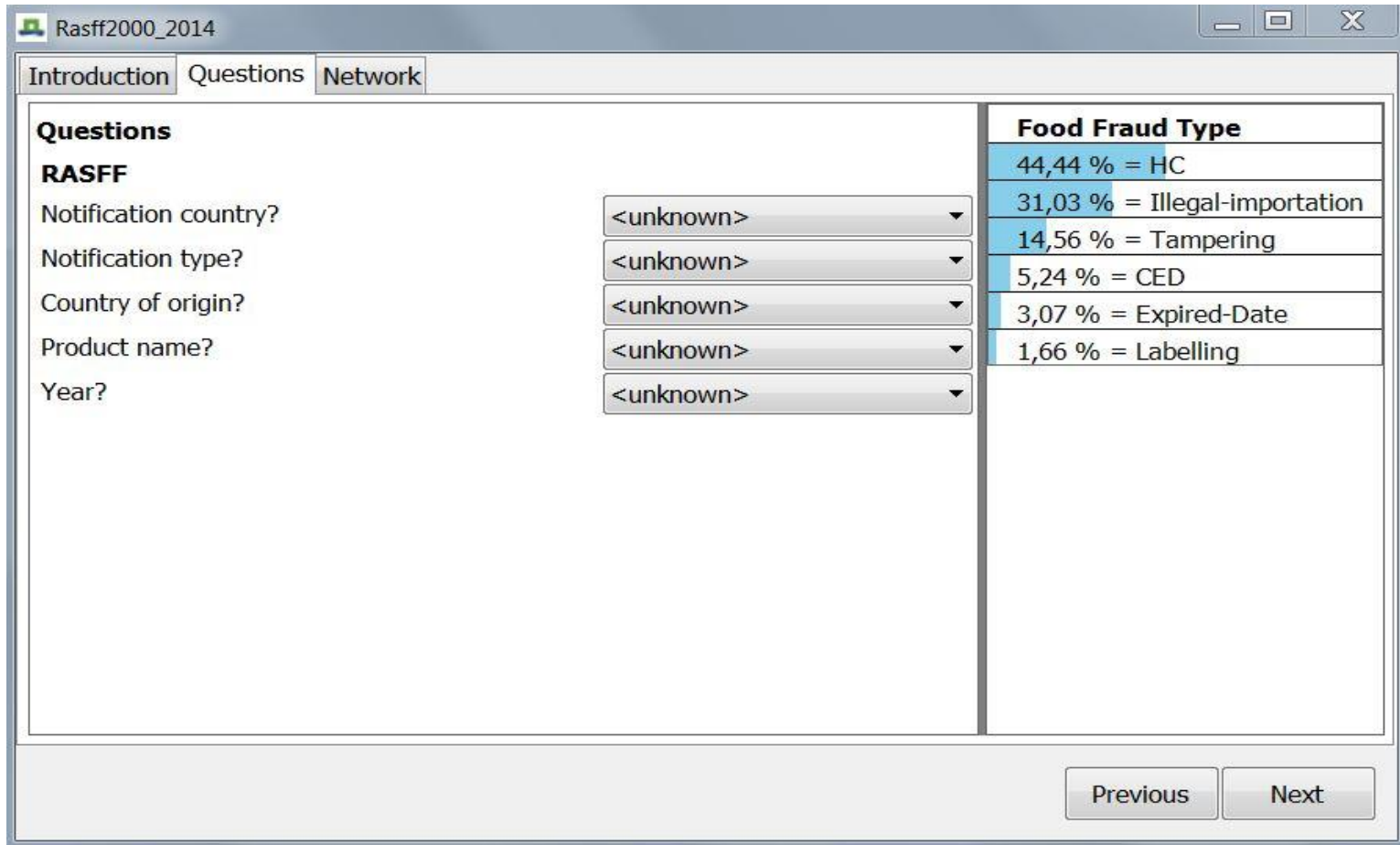
Modelling of Fraud in RASFF

Model developed with RASFF data up to 2013



Application of BNs to predict food fraud

USER-INTERFACE



The screenshot shows a software window titled "Rasff2000_2014" with three tabs: "Introduction", "Questions", and "Network". The "Questions" tab is active, displaying a questionnaire under the heading "RASFF". The questionnaire includes five questions, each with a dropdown menu currently set to "<unknown>".

Food Fraud Type
44,44 % = HC
31,03 % = Illegal-importation
14,56 % = Tampering
5,24 % = CED
3,07 % = Expired-Date
1,66 % = Labelling

At the bottom right of the window, there are two buttons: "Previous" and "Next".

Application of BNs to predict food fraud

CASE: What type of fraud can we expect from meat imported from Belgium

Rasff2000_2014

Introduction Questions Network

Questions

RASFF

Notification country? <unknown>

Notification type? <unknown>

Country of origin? Belgium

Product name? meat

Year? <unknown>

Food Fraud Type
71,84 % = Labelling
28,16 % = Tampering
0,00 % = CED
0,00 % = Illegal-importation
0,00 % = HC
0,00 % = Expired-Date

→ Labelling !

Previous Next



Application of BNs to predict food fraud

Validation of the Model;

Method:

Predict the type of fraud for all reports in RASFF in 2014.

Input values: type of product and country of origin

Results:

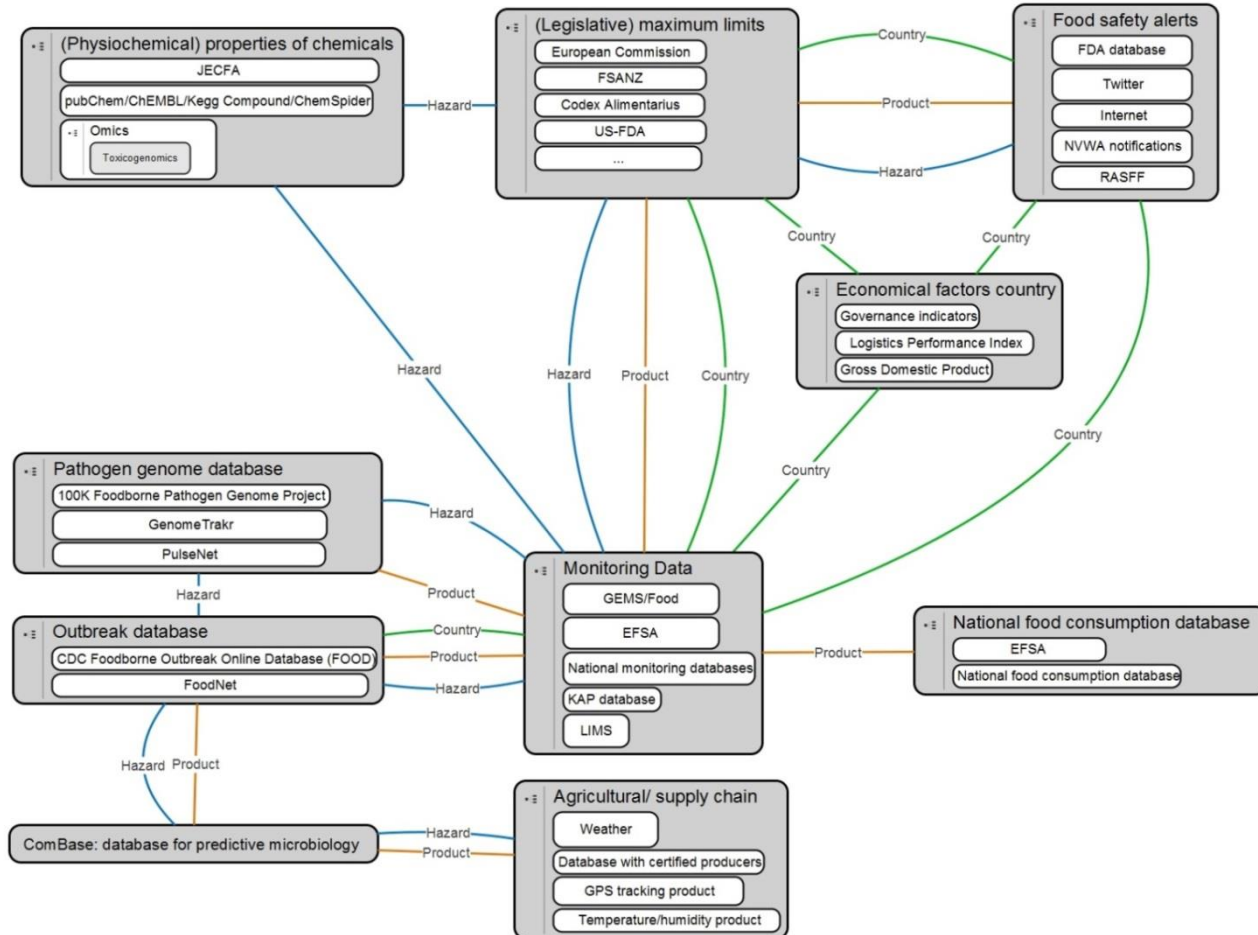
- 82% correct
- 51% prediction of new fraud combinations



Higher performance with dynamic model

Bayesian Networks (BNs) modelling in Big Data research

BN is able to combine data from different nature and location



Application of BNs to predict food fraud

Prediction of food fraud using multiple data sources and expert knowledge

- Identify parameters relevant for food fraud (\Rightarrow nodes in BN model)
- Determine the relationships between these parameters (expert knowledge, literature)
- Identify data sources of these parameters
- Develop the BN model
- Validate the BN model

Application of BNs to predict food fraud

Parameters relevant for food fraud (some examples)

Economic parameters:

- Prices of the fraudulent product at the time of detection
- Price spike around the period of detection
- Trade volumes of the product between the country of detection and country of origin
- Complexity of the food chain

Parameters of the country of origin & detection

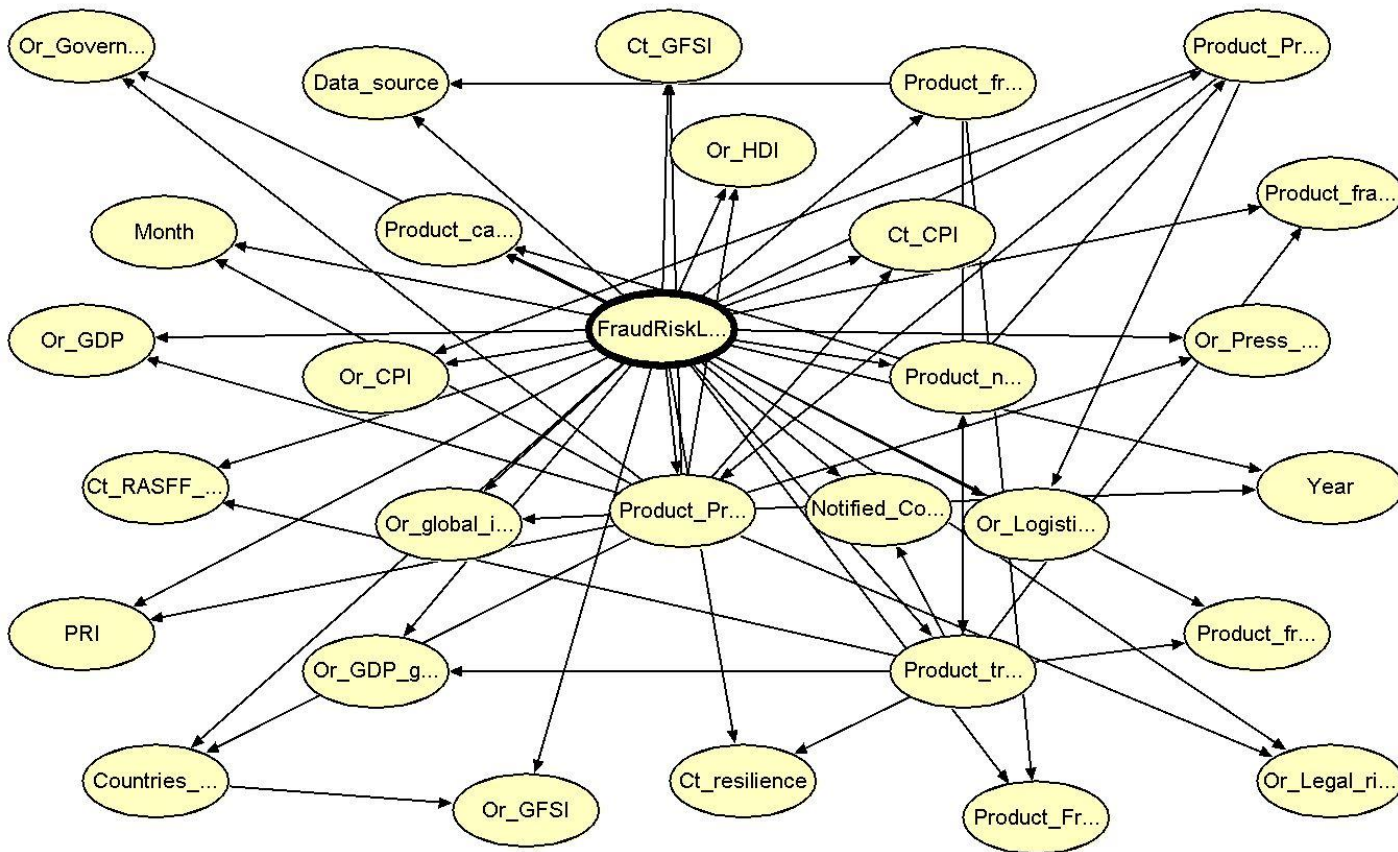
- Indices: corruption index, food safety index, governance index, legal system index, press index, human development index and technology index.

Detected food fraud cases

- RASFF, UPS and EMA

Application of BNs to predict food fraud

Preliminary BN model for predicting food fraud; relationships between all parameters



Preliminary BN model for predicting fraud linking 36 data sources (18 databases and 8 expert judgements)



Conclusions

- Big data developments show great potential in food safety research, food safety risk assessment and food safety risk management
- European Media Monitor MedISys is suitable to collect publications on food safety and food fraud on global level.
- Bayesian Network approach is useful to combine information from a variety of data sources
- Bayesian network models are useful to predict food fraud

Thank you

谢谢

