

PREDICTIVE ANALYSIS

Proof of Concept

Alessandro EL KHOURY Legal and Policy Officer DG SANTE – Unit G5



Scope of the exercise

- 1. Build a statistical model that uses data in IMSOC to predict a certain level of risk to commodities imported in the EU.
- 2. Use the results of point 1 to calculate frequencies of controls.







SYSTEMS AVAILABLE AND SYSTEMS USED

Available system components of **IMSOC:**

- TRACES
- RASFF
- AAC
- EUROPHYT
- ADNS

Systems with adequate data:

- TRACES
- RASFF
- AAC

System used:

TRACES

Key elements:

- Structure of data Pertinence of data
- Granularity of data
- Quantity of data

Food Safetv



SYSTEMS AVAILABLE AND SYSTEMS USED

RASFF VS TRACES

RASFF - For risks	TRACES - For certification
Variability of controls	Structured controls (import)
Limited to food and feed (mainly internal market)	All import commodities requiring a certificate / entry document
Data on risks	Customs data + controls
Less useful variables for the exercise	More useful variables for the exercise
Only negative results (when risk occurs)	Negative and positive results (good certificates and rejected certificates)



1. Categories of goods chosen

- Meat products (bovine, porcine and poultry)
- Different CN codes for meat products
- Grouping exercise



Initial Dataset



Why did we choose those 3 categories for this proof of concept?

Given the time-constraint and the limited scope of the exercise:

- Good number of CVED-Ps
- Relatively easier to create product categories
- Good number of rejections
- Comparable results with the same methodology





Initial Dataset



Accepted and rejected

2011-today

Food Safety



Initial Dataset Assign a score

EUI	ROPEAN UNION				The Common Veterinary	y Entry Document, CVED
	24. Previous CVED:		No	Yes	25. CVED Reference Number:	Local reference number
	Reference number:				CVEDP.BE.2018.0007316 - V1	803166
	26. Documentary Check:				27. Identity Check:	
T	Satisfactory	\sim	Not satisfactory		Seal check	OR Full identity check
e					Satisfactory	Not satisfactory
E						
5						
S	28. Physical Check:				29. Laboratory Tests: No 🔀	Yes
consignment						
õ	Satisfactory		Not satisfactory		Tested for::	
E	Not done				Random Suspicion	Re-enforced
2	1. Reduced checks regime			\sim	Results:: Satisfactory	Not satisfactory
5	-				-	
	2. Other				Released pending a result	
. <u>.</u>						
decision on	30. ACCEPTABLE for Transhipmen	t			31. ACCEPTABLE for Transit Procedure	
	_					
Part II:	EU BIP		TRACES unit no.		To 3rd Country	+ ISO code
t						
a	3rd country		3rd Country ISO code		Exit BIP	TRACES unit no.
-						
	32. ACCEPTABLE for Internal Mark	(et			33. ACCEPTABLE if channelled	
	For Free Circulation				Article 8 procedure	
	Human consumption:		\sim		Re-import of EU products (Article 15)	
			\simeq		Re-import of EO products (Africie 15)	
	Animal feedingstuff:					
	Pharmaceutical use:				34. ACCEPTABLE for Specific Warehouse Proces	dure(Articles 12.4 and 13)
	Technical use:					
	Other:				Customs warehouse	
					Free zone or Free warehouse	
	35. NOT ACCEPTABLE				Ship supplier	
					Direct te a ship	
	1. Re-export					
	2. Destruction				36. Reason for Refusal	
	3. Transformation					
	By Date:				1. Absence/Invalid certificate	
	by Date:					
					2. Non approved country	
	37. Details of Controlled Destinations	s (33-35)			Non approved establishment	
	Approval no (where relevant)				4. Prohibited product	
	Address		67	7	5. ID: Mis-match with documents	
				7	D: Health mark error	
					7. Physical hygiene failure	
					8. Chemical contamination	
	38. Consignment Resealed				9. Micro biological contamination	
	New seal no: ktn 08	377937	1		10. Other	

Accepted/Rejected	Score
Accepted certificate	0
Absence/invalid certificate	1
Non approved country	1
Non approved establishment	1
Prohibited product	7
ID mis-match with documents	3
ID Health mark error	5
Physical hygiene failure	7
Chemical contamination	10
Microbiological contamination	10
Other	10



Initial dataset looks like...

	N COMMU	UNITY				The Con	nmon Ve	terinar	y Ent	ry Doc	ument	t, CVEI
1. Consignor / E	Doporter					2. CVED refere	nce number		Local r	eference numb	er:	
Name	-		0			-	0			Ŧ	0	
Address	-		0			Border Inspection Post						
	Ŧ		0			-			0			
						TRACES Unit 1	lumber :					
Country	Ŧ		0			-	0					
2.0												
3. Consignee	Ŧ		0			4. Person respor	sible for load			0		
Name						Name						
Address	-		0			Address	-			0		
			0			1	-			0		
Country	Ŧ		0			Country	-			0		
5. Importer						6. Country of or		+ ISO 00	de 7 Cour	try from when	e consigned	+ ISO oc
Name	Ŧ		0						_			1.00 0
Address	-		0			8. Delivery addr				- Argentin	а јак	
2000.000	w		0			o. Denvery adu	*			0		
						ʻ	*			0		
										0		
Country	T		_									
9. Arrival at BIF	P (estimated date)	▼ □)			10. Veterinary d Number(s)						
11. Vanal name	o / Trlight Ma						*			0		
11. Vessel name / Flight No. Aeroplane Ship Railway waqon												
Aen	oplane	Ship		Railway w	agon	Date of issue	-					
	oplane	Ship		Railway w	/aqon ▶	Establishment o	f origin(where rele	evant)		-	0	▶ 8
Identification::		Ship)	raqon ⊫		f origin(where rele			•	0	▶ @
⊲ Identification:: Documentary re	eferences:)	raqon ►	Establishment o	f origin(where rele	evant) +		-	0	
Identification: Documentary re 16. Seal number	eferences:)	/aqon ▶	Establishment o	f origin(where rele oval number				0	
⊲ Identification:: Documentary re	eferences:)	raqon ►	Establishment o	f origin(where rele					
Identification: Documentary re 16. Seal number	eferences:)	raqon ►	Establishment o	f origin(where rele oval number					
Identification: Documentary re 16. Seal number Seal Num	ferences:	-)	raqon ►	Establishment o	f origin(where relioval number					
Identification: Documentary re 16. Seal number	ferences:)	iaqon	Establishment o	f origin(where rele oval number					
Identification: Documentary re 16. Seal number Seal Num	ferences:	-)	iaqon	Establishment o	f origin(where relioval number					
Tentification:: Documentary re 16: Seal number Seal Num Container	ferences:	-)	iaqon	Establishment o Veterinary appr	f origin(where rel oval number					
Identification:: Documentary re 16: Seal number Seal Num Container	ferences:	*			raqon ▶	Establishment o Veterinary appr 18. For transit to	6 crigin(where relined in the second	*	No	Yes		
Tentification:: Documentary re 16: Seal number Seal Num Container	ferences:	*)	iaqon ia	Establishment o Veterinary appr	f origin(where rel oval number	*				
Identification:: Documentary re 16: Seal number Seal Num Container	ferences:	*			iagon	Establishment o Veterinary appr 18. For transit to	6 crigin(where relined in the second	*		Yes		
Identification: Documentary re Seal number Seal Num Container 17. Transhipme EU BIP	ferences:	*	Yes		iagon	Establishment o Veterinary appr 18 For transit to To 3rd Country	f origin(where reb or al number	*	No_	Yes		
Identification: Documentary re 16 Seal number Seal Num Container T. Transhipme EU BIP Ide country [19. Conform to: Doce NOL C	ferences:	*	Yes		raqon	Establishment o Veterinary appr 18. For transit to To 3rd Country Exit BIP.	f origin(where reb or al number	*	No_	Yes		
Identification: Documentary re 16 Seal number Seal Num Container Transhipme EU BIP 3rd country 19. Conform to:	ferences:	*	Yes		iagon	Establishment o Veterinary appr 18. For transit to To 3rd Country Exit BIP.	f origin(where reb or al number	*	<u>No</u>	- Yes		
Identification: Documentary re Seal number Container T. Transhipme EU BIP Ide country [Contorn to: DeceNDLC Context	ferences:	*	Yes		iaqon	Establishment o Veterinary appr 18: For transit to To 3rd Country Boxt BIP: 20: For re-impos	f origin(where reb or al number	-	<u>No</u>	- Yes		
Identification: Identification: Documentary re Seal number Seal Numi Container 17. Transhipme 19. Contern to: 19. Conform to: Does NOT conformer Conforms	ferences:	*	Yes		aqon B	Establishment o Veterinary appr 18: For transit to To 3rd Country Boxt BIP: 20: For re-impos	Corigin(where rel or al number)))))))))))))))))))	-	<u>No</u>	Yes 0		
Identification: Identification: Documentary re Seal Numi Container 10: Seal Numi Container 17: Transhipme EU BIP Ide country 19: Conform to: Does NOT c Conternal Feec Animal Feec Human Con	ferences:	*	Yes		raqon B	Establishment o Veterinary approved 18. For transit to To 3rd Country Exit BIP: 20. For re-impose 22. For NON-C Oustorns We Direct To A	c origin(where ref oval number)))))))))))))))))))	•	<u>No</u>	Yes 0		
Gentral Section (Contention) Gentral Section (Container Seal Num Container Container T. Transhipme EU BIP If Conform to: The section of the	ferences:	*	Yes		aqon	Its For transit to Veterinary approved to To 3rd Country Exit BIP. 20. For re-impo Customs We Direct To A Free Zone C	Corigin(where ref oval number)) 3rd Country)) 3rd Country)) 1 t t t t t t t t t t t t t t t t t	•	<u>No</u>	Yes Person Person Pers		
Identification: Identification: Documentary re Seal Numi Container 10: Seal Numi Container 17: Transhipme EU BIP Ide country 19: Conform to: Does NOT c Conternal Feec Animal Feec Human Con	ferences:	*	Yes		aqon B	Establishment o Veterinary approved 18. For transit to To 3rd Country Exit BIP: 20. For re-impose 22. For NON-C Oustorns We Direct To A	Corigin(where ref oval number)) 3rd Country)) 3rd Country)) 1 t t t t t t t t t t t t t t t t t	•	<u>No</u>	Yes Person Person Pers		

Each field in the certificate corresponds to a **variable** For each certificate there are different variables. Technically this can be pictured as an excel spreadsheet...



- Lines (horizontal) represents a certificate
- Columns (vertical) represents the different variables of the certificate

Consignor/Exporter $ riangle$	Consignor Address	Consignor Post Code	Consignor City	Consignor Country	Country ISO Code	CVED Reference Number	Local Reference Number	Border Inspection Post	Entry BIP Country	Country ISO Code	TRACES Unit Number	Consignee
Total												
	Scalabrini Ortiz 215 Piso 3 Departamento B	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.ES.2018.0010626	1VLC1-7900609764013	Valencia, P	Spain	ES	ESVLC1	REINA APICOLA LEVANTINA, S.L.
	Scalabrini Ortiz 215 Piso 3 Departamento B	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.ES.2018.0012474	1VLC1-7900609764004	Valencia, P	Spain	ES	ESVLC1	REINA APICOLA LEVANTINA, S.L.
BM SUPPLIER SRL	BLANCO ENCALADA 197 PB OF 49 SAN ISIDRO	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.MT.2018.0000935	2018/0892	Marsaxlokk, P	Malta	MT	MTMAR1	EURL SIRINE TRADING
CASAMEN S.A.	MAXIMO PAZ735 Y 963 (B1824KSO) LANUS OESTE	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.ES.2018.0004850	1VLC1	Valencia, P	Spain	ES	ESVLC1	ALBERTO L. HUGHES, S.L.
	Comandante Franco 4901	Buenos Aires	Bernal	Argentina	AR	CVEDP.DE.2018.0013240	100-491-18	Hamburg Hafen	Germany	DE	DEHAM1	Block House Fleischerei GmbH
Congeladores Patagonicos 5.A.		Chubut	Puerto Madryn	Argentina	AR	CVEDP.ES.2017.0061261	1MAR1-7900603609236	Marín, P	Spain	ES	ESMAR1	PESCANIGRAN S.L.
Congeladores Patagonicos 5.A.		Chubut	Puerto Madryn	Argentina	AR	CVEDP.ES.2017.0061272	1MAR1-7900603609281	Marín, P	Spain	ES	ESMAR1	PESCANIGRAN S.L.
DARSENA DE	PUERTO MAR DEL PLATA 7600 MAR DEL PLAT	Montevideo	Montevideo	Uruguay	UY	CVEDP.NL.2018.0016966	18033688	Rotterdam	Netherlands	NL	NLRTM1	/
	COLOMBIA N81005					Health and						LANDALIED





Data Manipulation phase

Data was "cleaned" \rightarrow manipulated to make it simpler for machine learning algorithm to use

- 1. Delete certificates with multiple CN codes (<1%)
- 2. delete variables with >20% missing values
- 3. Delete all columns that cannot be converted into numbers (i.e. name of operators)
- 4. Delete geographical data (multiple methods possible, currently outside the scope of the exercise)
- Replace missing values for variables with <20% (prediction model)





Commission

Final dataset

- Divided in 2 parts (randomly generated):
- 80% to learn
- 20% to predict

		Total
Bovine - Train	634	793
Bovine - Test	159	795
Porcine - Train	253	316
Porcine - Test	63	510
Poultry - Train	1110	1387
Poltry - Test	277	1307
Total observations after cleaning		2496

3 different categories

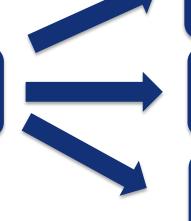
Bovine meat

Porcine meat

Poultry meat



1 model (random forest)



Food Safety



Predictive model's performance

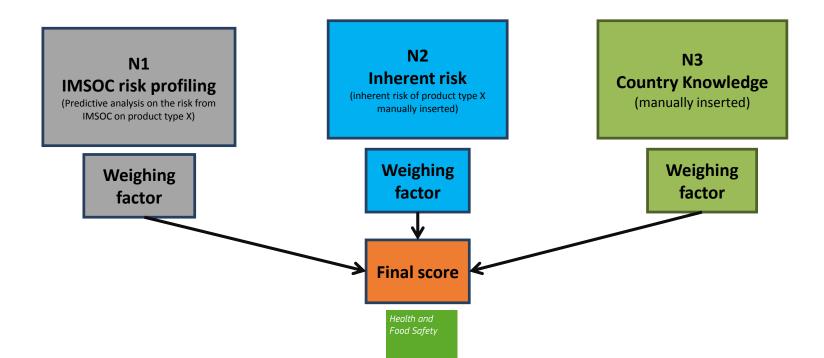
	Bovine	Porcine	Poultry
Absence of non compliances	100%	~80%	100%
Documentary checks fail	~60%	~70%	~50%
Identity checks fail	~40%	-	~60-70%
Physical checks fail	~80%	~60%	~95%





Possible mathematical approach to calculate the frequency

N1, N2, N3 would give 3 sub-scores between 0 and 1, and a weighing factor is applied to each sub-score.





Two possible methods to calculate the frequency

Method A: sum <u>(frequency per consignment)</u> N1+N2+N3 -> algorithm to calculate the frequency

Method B: multiplication <u>(frequency per product</u> <u>category)</u> N1*N2*N3 -> algorithm to calculate the frequency





METOD A: SUM (Frequency per individual consignment)

Weighed N1, N2 and N3 are added to determine a final score for each individual consignment.

- N1 varies for each consignment, according to the likelihood of rejection;
- N2 and N3 are stable;

This final score is used in an algorithm to calculate the frequency for each individual consignment.

Result: different frequency for each individual consignment of the same product type.





METOD B: MULTIPLICATION (Frequency per product category)

N1, N3 and N3 have the same weight.

- N1 is given by the average of the reasons of rejection for a specific product category.
- N2 and N3 are stable.

N1*N2*N3 are used in an algorithm to calculate a frequency for a specific product category.

Result: one common frequency for all consignments of the same product type coming from a specific third country.





QUESTIONS

- 1. Do we need to have a separate frequency for Identity checks?
- 2. Do you prefer a frequency established per each consignment, or one per product category?





Default 100% frequencies

- Animals
- Plant and plant products
- Certain POAO and ABP

Possibility to lower the default rate if needed.





Thank you for the attention!

SANTE Unit G5 – Alerts, Traceability and Committees TRACES-IMSOC sector SANTE-IMSOC@ec.europa.eu