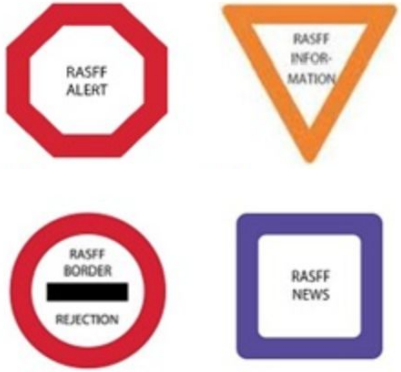
AI meets Food Safety 🥗

Predicting Food Safety Risk with Machine Learning

Rapid Alert System for Food and Feed (RASFF)



ORIGIN	PROCESS		MEASURES	
CASE STUDY: E.coli 	**Notifications on E.coli** food-borne outbreak linked to fresh vegetables (05/2011 - 07/2011)	**Tracing back foods** linked to the outbreak to find "hotspots". Once the contaminated food is identified: tracing the products in which it is present	RASFF ALERT Fenugreek seeds from Egypt	Fenugreek seeds, mixes of seeds, sprouts withdrawn in EU27 countries

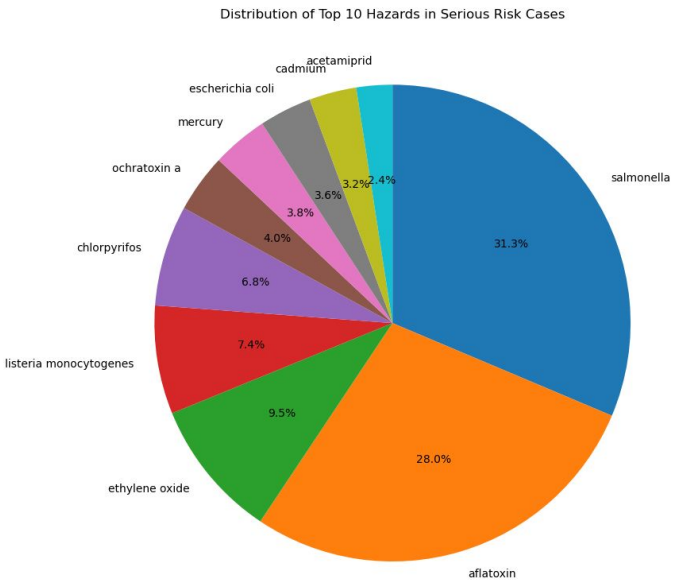
RASFF dataset (2020-2025)

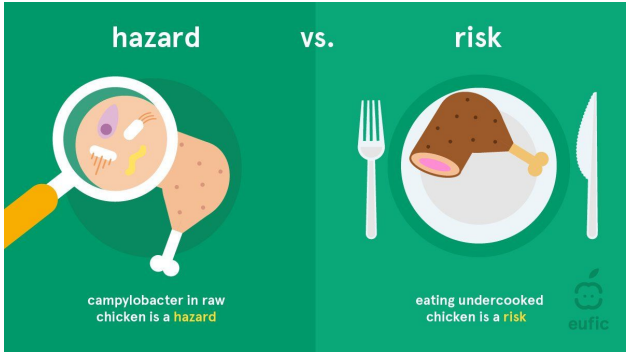
14 features
27398 cases

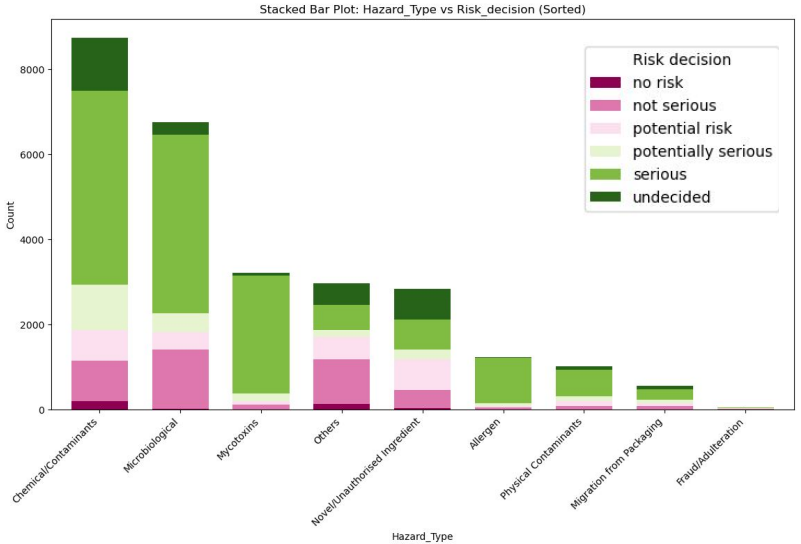
Data cleaning

Category	**37 categories**
Type:	**6 types (food, feed etc)**
Subject :	**complex text (English + EU languages)** e.g . suspected Salmonella i beef and hamburgers from Sweden
Hazard	**1641 categories**
Origin	**609 categories with multiple origin**
Notifying country	**33 categories**
Operator	**4876 categories with multiple origin**
Distribution	**4793 categories with multiple origin**
Classification	**4 main categories**
Risk decision	**6 risk categories**

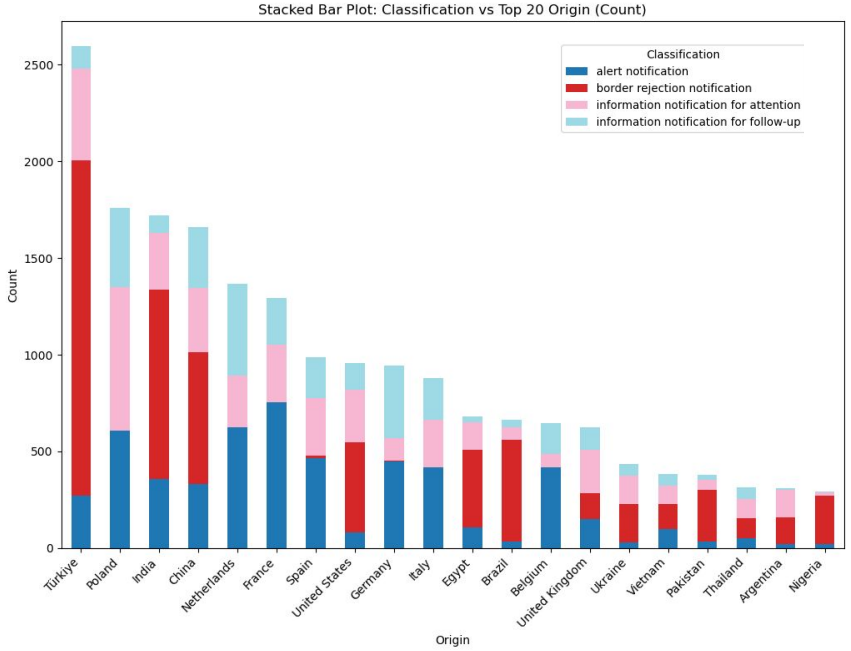
Hazard





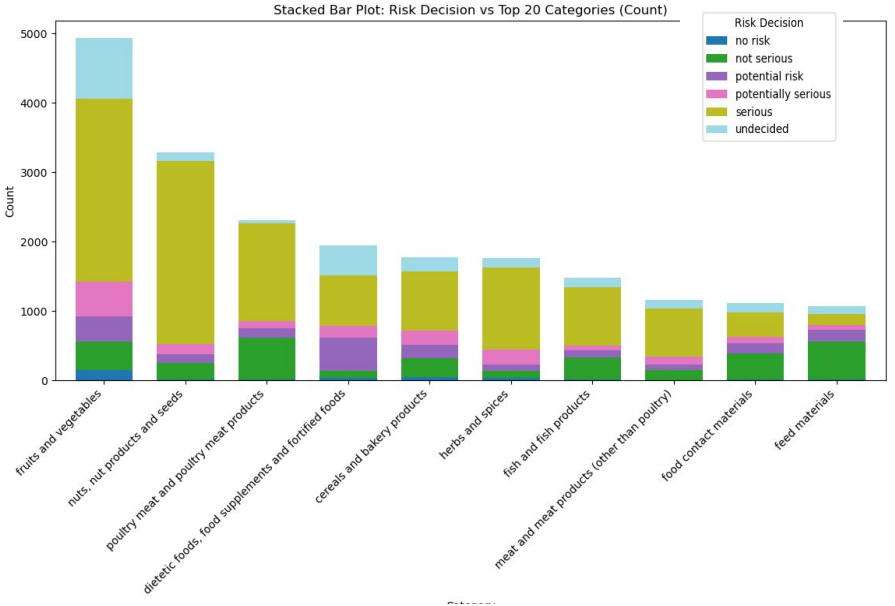


9 different hazard type were defined based on the subject text using LLMs



Origin

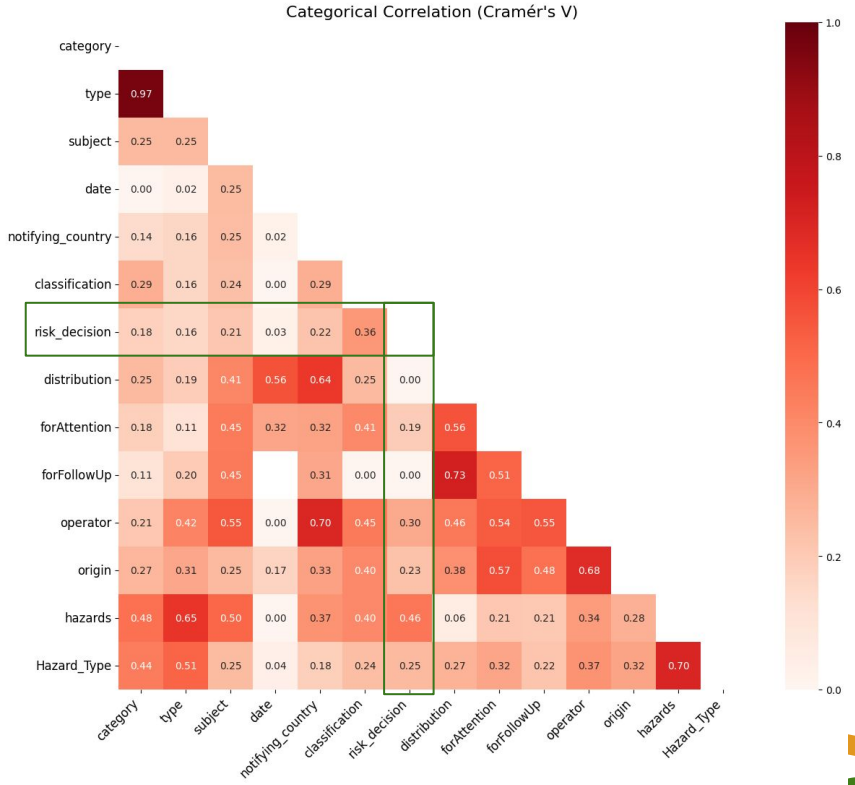
Category



Feature selection

7 Features :
Category, subject, origin, notifying country, hazard, hazard type, classification

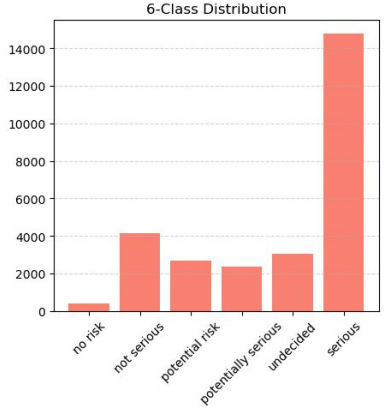
Target:
Risk decision (6 levels)

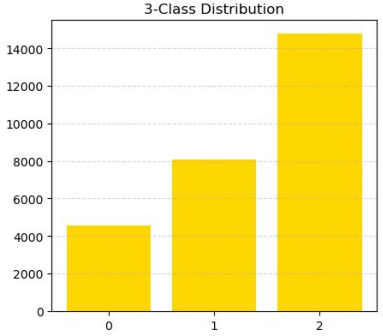


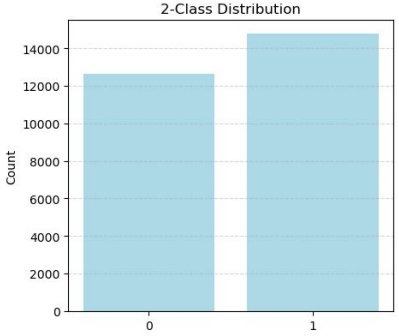
Feature engineering

Features	Properties	Engineering
Category	37 categories	Sentence Transformer embedding or Group rare categories + One-hot encoding
Subject :	complex text	Sentence Transformer embedding using 'all-MiniM-L6-v2'
Hazard	1641 categories	Target encoding + smoothing
Origin	609 categories with multiple origin	Target encoding + smoothing
Notifying country	33 categories	Target encoding + smoothing
Hazard Type	9 categories	One-hot encoding
Classification	4 main categories	One-hot encoding

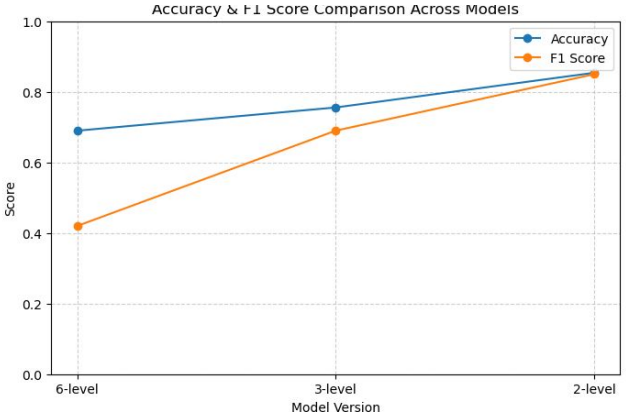
Main challenge 1: class imbalance

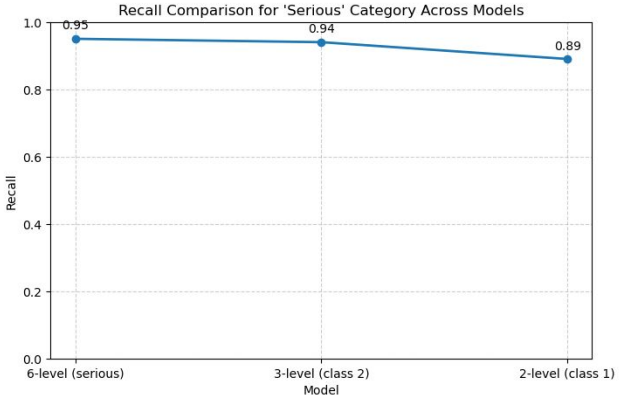




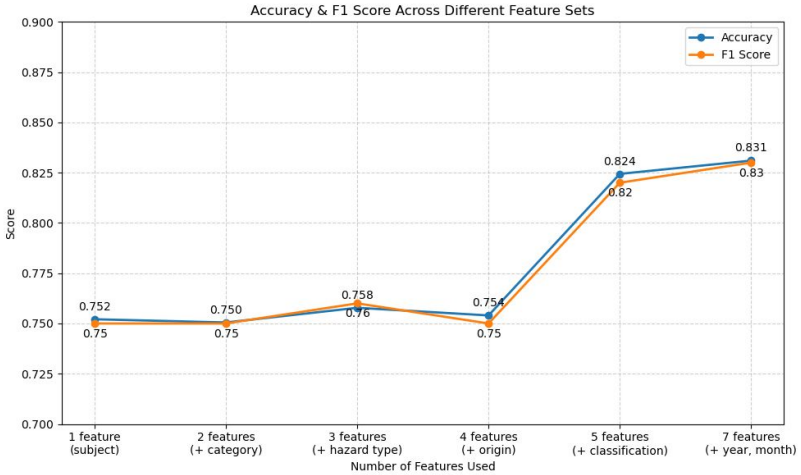


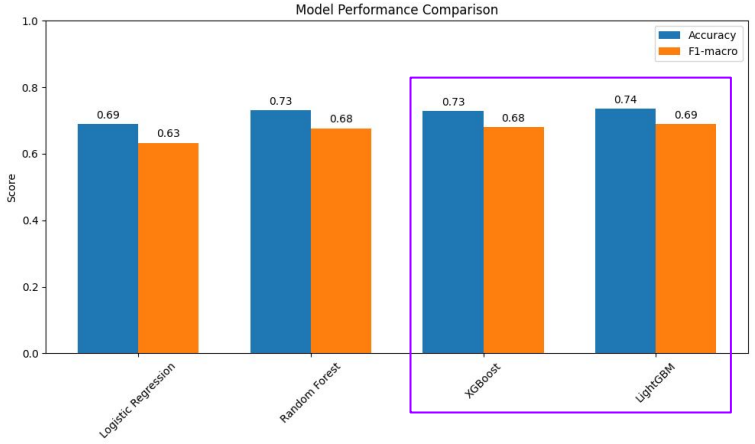
- One-hot encoding
- XGboost





Main challenge 2: High carlinarlity

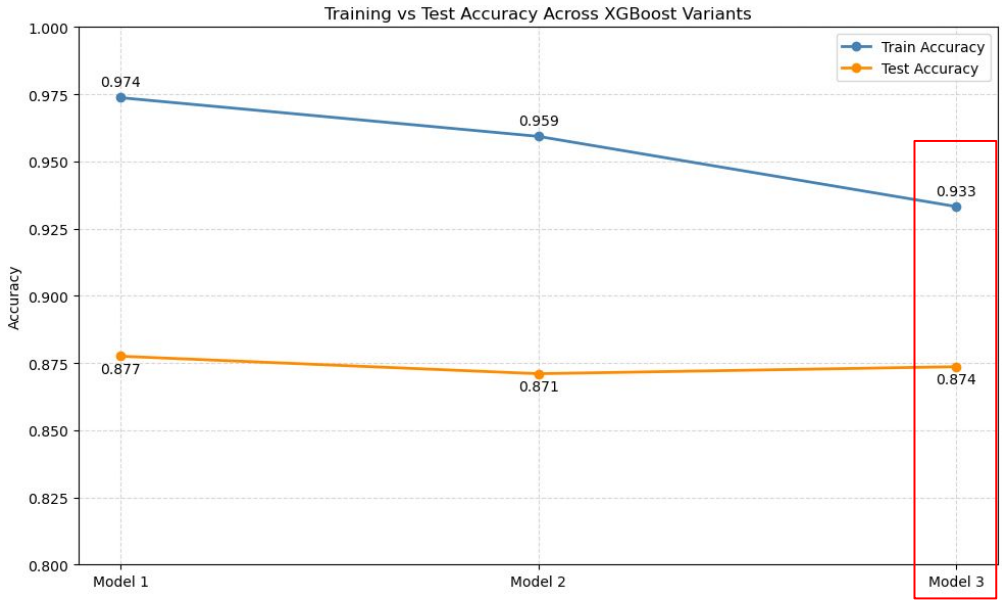




- 3 level classification
- XGboost

- Model comparison in 4 features sets

Model Tuning & Best Model



XGBoost best model parameter	
n-estimator	500
max_dept	4
learning_rate	0.05
subsample	0.8
colsample_bytree	0.8
gamma	0.1
min_child_weight	1
reg_alpha	1
reg_lambda	2

Accuracy: 0.87
F1 score: 0.87

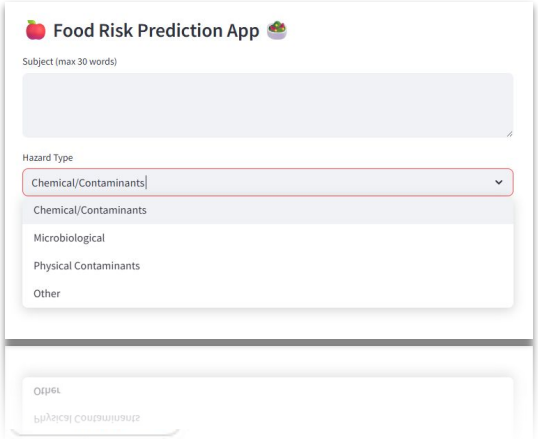
Precision
Risk: 0.91
No-Risk: 0.89

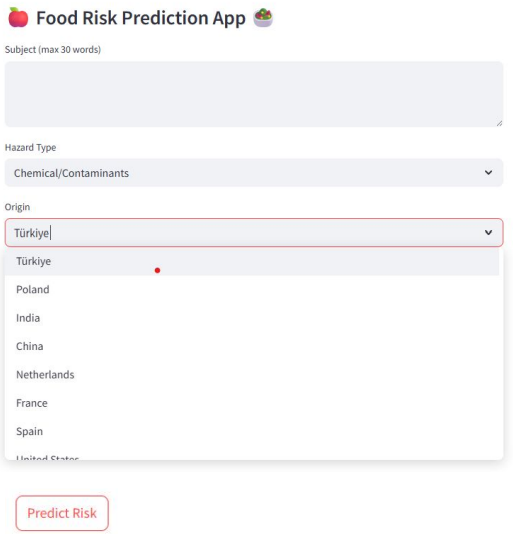
Recall
Risk: 0.91
No-Risk: 0.83

Demo streamlit

User input
- Subject (text)
- Hazard Type (select)
- Origin (select)

Output
- Risk probability





⚠️ High Risk AlertThis notification is predicted to belong to the high risk group.Predicted probability of serious risk: 59.86%.It has a potentially high risk that could cause issues. Please take action accordingly.

✅ Moderate/Low RiskThis notification is predicted to belong to the medium to low risk group.Predicted probability of serious risk: 33.00%.Stay alert and follow up as needed.

Summary & Future Work

- Create the risk predictive model (87%) with 7 features
- Improvement on the model performance for real-world application
  e.g. feature engineering (ideally more features), data enrichment using fine-tuned transfer models
- Improved version of application
- Demonstrate the potential of using ML to food safety management system

THANKS

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**