Food Hazard Detection

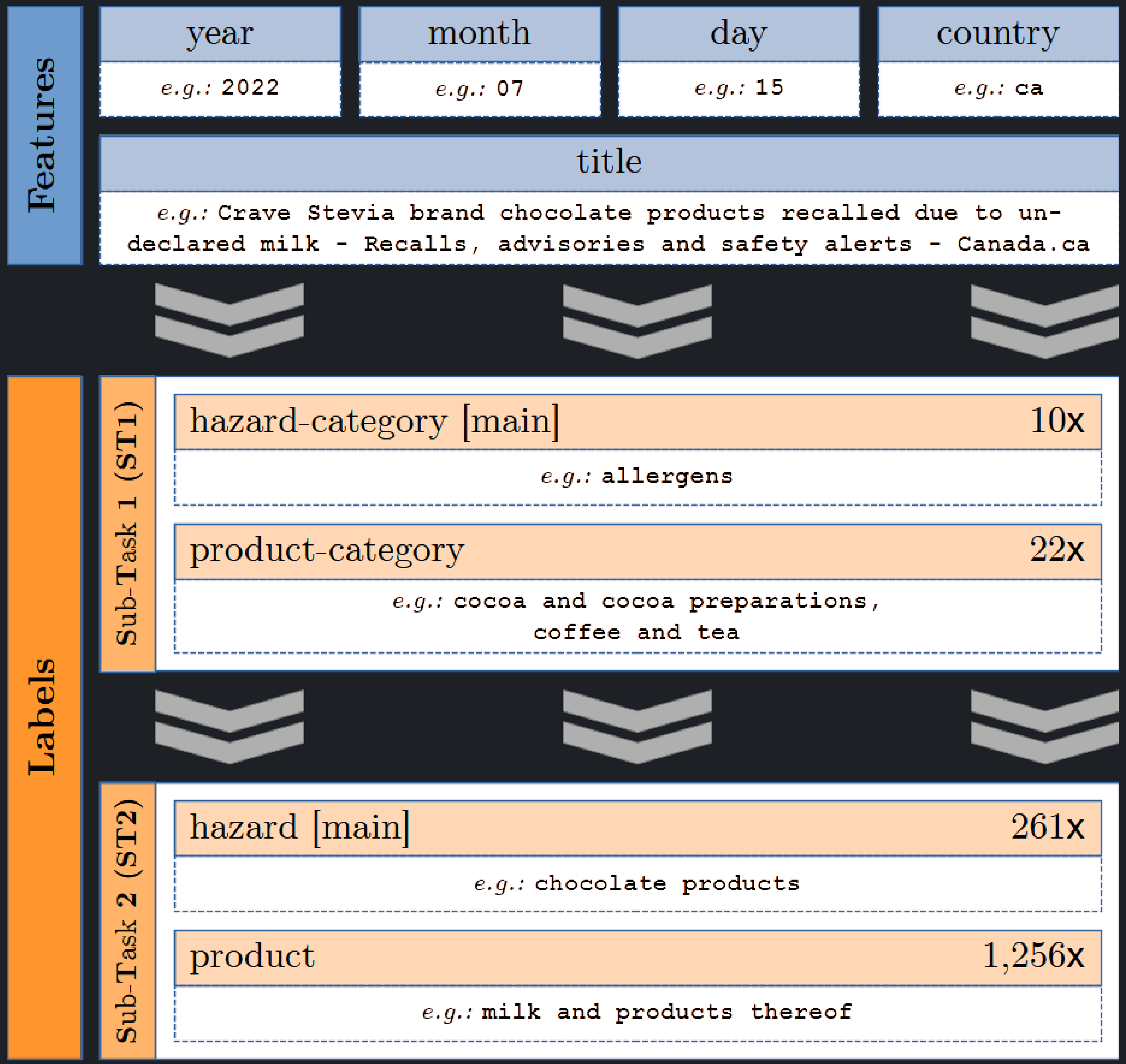
Introduction to Deep Learning

Meesum Abbas
Muhammad Saad

- SemEval Project
- Given incident reports
- Total train dataset - 5,082

Sub-tasks

- **ST1 - Category classification** - Predicting the type of hazard and product.
- **ST2 - Item Classification** - Predicting the exact hazard and product


Features
year
e.g.: 2022
month
e.g.: 07
day
e.g.: 15
country
e.g.: ca
title
e.g.: Crave Stevia brand chocolate products recalled due to un-declared milk - Recalls, advisories and safety alerts - Canada.ca
Labels
Sub-Task 1 (ST1)
hazard-category [main]
10x
e.g.: allergens
product-category
22x
e.g.: cocoa and cocoa preparations, coffee and tea
Sub-Task 2 (ST2)
hazard [main]
261x
e.g.: chocolate products
product
1,256x
e.g.: milk and products thereof

Evaluation

```python
from sklearn.metrics import f1_score

def compute_score(hazards_true, products_true, hazards_pred, products_pred):
  # compute f1 for hazards:
  f1_hazards = f1_score(
    hazards_true,
    hazards_pred,
    average='macro'
  )

  # compute f1 for products:
  f1_products = f1_score(
    products_true[hazards_pred == hazards_true],
    products_pred[hazards_pred == hazards_true],
    average='macro'
  )

  return (f1_hazards + f1_products) / 2.
```

Problem Statement

Identify food hazard/product categories, and extract precise hazard/product information from diverse and unstructured web-sourced reports

Models, Experiments, Explainers
ST1

Existing Approaches

- Classification models
- BERT
- Variants

BERT - Baseline

Hazard-Category

	precision	recall	f1-score	support
allergens	0.89	0.90	0.90	377
biological	0.86	0.92	0.89	339
chemical	0.78	0.62	0.69	68
food additives and flavourings	0.00	0.00	0.00	5
foreign bodies	0.70	0.84	0.77	111
fraud	0.81	0.56	0.66	68
migration	0.00	0.00	0.00	1
organoleptic aspects	0.50	0.30	0.38	10
other hazard	0.55	0.44	0.49	27
packaging defect	0.38	0.27	0.32	11
accuracy			0.83	1017
macro avg	0.55	0.49	0.51	1017
weighted avg	0.82	0.83	0.82	1017

BERT - Baseline

Product-Category

	precision	recall	f1-score	support
alcoholic beverages	0.71	0.71	0.71	7
cereals and bakery products	0.73	0.86	0.79	123
cocoa and cocoa preparations, coffee and tea	0.77	0.73	0.75	49
confectionery	0.58	0.53	0.55	40
dietetic foods, food supplements, fortified foods	0.61	0.71	0.65	24
fats and oils	0.50	0.50	0.50	4
feed materials	0.00	0.00	0.00	3
food contact materials	0.00	0.00	0.00	1
fruits and vegetables	0.81	0.79	0.80	112
herbs and spices	0.54	0.44	0.48	16
honey and royal jelly	0.00	0.00	0.00	1
ices and desserts	0.92	0.88	0.90	56
meat, egg and dairy products	0.90	0.93	0.91	282
non-alcoholic beverages	0.96	0.84	0.90	31
nuts, nut products and seeds	0.76	0.81	0.78	63
other food product / mixed	1.00	0.11	0.20	9
pet feed	0.00	0.00	0.00	6
prepared dishes and snacks	0.56	0.49	0.52	90
seafood	0.87	0.95	0.91	56
soups, broths, sauces and condiments	0.66	0.72	0.69	43
sugars and syrups	0.00	0.00	0.00	1
accuracy			0.79	1017
macro avg	0.57	0.52	0.53	1017
weighted avg	0.78	0.79	0.78	1017

Early Stopping

Classification Report for ST1 – Hazard Category with Early Stopping:

	precision	recall	f1-score	support
allergens	0.82	0.92	0.87	186
biological	0.93	0.89	0.91	175
chemical	0.90	0.66	0.76	29
food additives and flavourings	0.00	0.00	0.00	2
foreign bodies	0.81	0.77	0.79	56
fraud	0.69	0.65	0.67	37
organoleptic aspects	0.00	0.00	0.00	5
other hazard	0.50	0.64	0.56	14
packaging defect	0.60	0.60	0.60	5
accuracy			0.84	509
macro avg	0.58	0.57	0.57	509
weighted avg	0.83	0.84	0.83	509

Class Weights

Classification Report for ST1 - Hazard Category:

	precision	recall	f1-score	support
allergens	0.88	0.85	0.87	371
biological	0.90	0.90	0.90	348
chemical	0.68	0.72	0.70	57
food additives and flavourings	0.43	0.60	0.50	5
foreign bodies	0.81	0.79	0.80	112
fraud	0.63	0.65	0.64	74
migration	0.00	0.00	0.00	1
organoleptic aspects	0.29	0.36	0.32	11
other hazard	0.47	0.52	0.49	27
packaging defect	0.54	0.64	0.58	11
accuracy			0.82	1017
macro avg	0.56	0.60	0.58	1017
weighted avg	0.82	0.82	0.82	1017

Learning Rate Scheduling

Classification Report:

	precision	recall	f1-score	support
allergens	0.85	0.90	0.87	186
biological	0.89	0.92	0.91	175
chemical	0.74	0.69	0.71	29
food additives and flavourings	0.00	0.00	0.00	2
foreign bodies	0.82	0.75	0.79	56
fraud	0.66	0.68	0.67	37
organoleptic aspects	1.00	0.60	0.75	5
other hazard	0.69	0.64	0.67	14
packaging defect	0.00	0.00	0.00	5
accuracy			0.84	509
macro avg	0.63	0.58	0.60	509
weighted avg	0.83	0.84	0.83	509

Augmentation using Synonyms

Classification Report:

	precision	recall	f1-score	support
allergens	0.95	0.92	0.93	286
biological	0.96	0.96	0.96	275
chemical	0.98	0.94	0.96	129
food additives and flavourings	0.99	1.00	1.00	102
foreign bodies	0.86	0.96	0.91	157
fraud	0.92	0.93	0.93	137
migration	1.00	1.00	1.00	100
organoleptic aspects	1.00	0.99	1.00	105
other hazard	1.00	0.95	0.97	113
packaging defect	0.98	0.99	0.99	105
accuracy			0.96	1509
macro avg	0.96	0.96	0.96	1509
weighted avg	0.96	0.96	0.96	1509

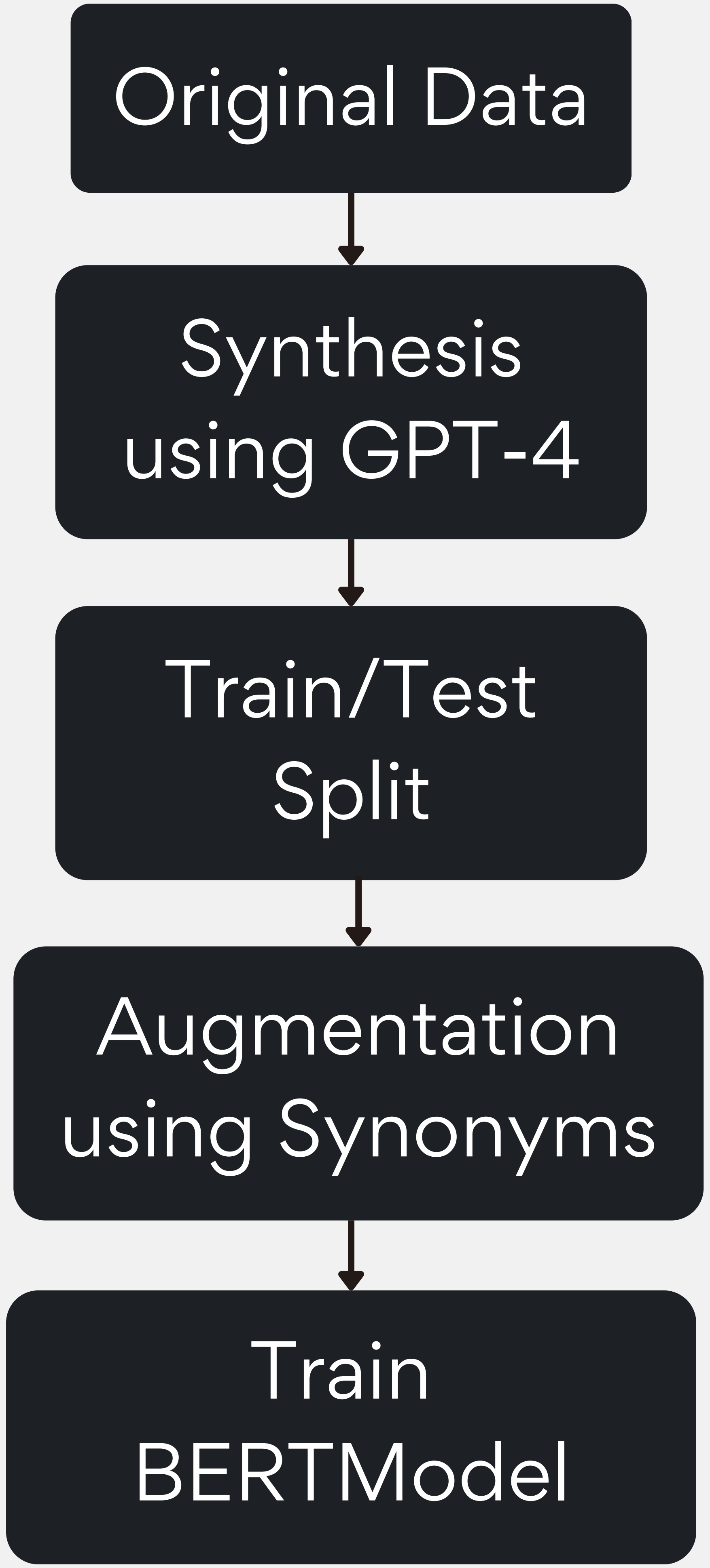
Result on SemEval

steliosg23	24	11/26/24		0.7060 (16)	0.0000 (1)	View
msaadg	5	11/06/24		0.7006 (17)	0.0000 (1)	View
Alex_goodman	7	11/06/24		0.6944 (18)	0.0000 (1)	View
meesuma5	2	11/24/24		0.6907 (19)	0.0000 (1)	View
BenPhan	9	11/20/24	Ryder	0.6877 (20)	0.0000 (1)	View

Correction

- Synthesis for minority classes
- Train/Test Split
- Augmentation
- Threshold - 600. Num_Augmentations - 800

Task Pipeline



Synthesis + Augmentation

	precision	recall	f1-score	support
allergens	0.94	0.98	0.96	341
biological	0.97	0.99	0.98	352
chemical	0.90	0.99	0.94	71
food additives and flavourings	0.83	0.83	0.83	6
foreign bodies	0.97	0.97	0.97	112
fraud	0.81	0.68	0.74	77
migration	1.00	1.00	1.00	7
organoleptic aspects	0.91	0.83	0.87	12
other hazard	0.81	0.69	0.75	32
packaging defect	1.00	0.68	0.81	22
accuracy			0.94	1032
macro avg	0.92	0.86	0.89	1032
weighted avg	0.94	0.94	0.94	1032

Comparison

- [5] [2022] - BERTSEQ 0.597
- [5] [2022] - BERTARE 0.491
- [2] [2023] - BERT 0.85

Future Directions

- **Ensembling** - Training 2 BERTs and Averaging
- More **Data Synthesis** - T5

Models, Experiments, Explainers
ST2

Class Distribution

- 128 Hazards
- 1022 Products

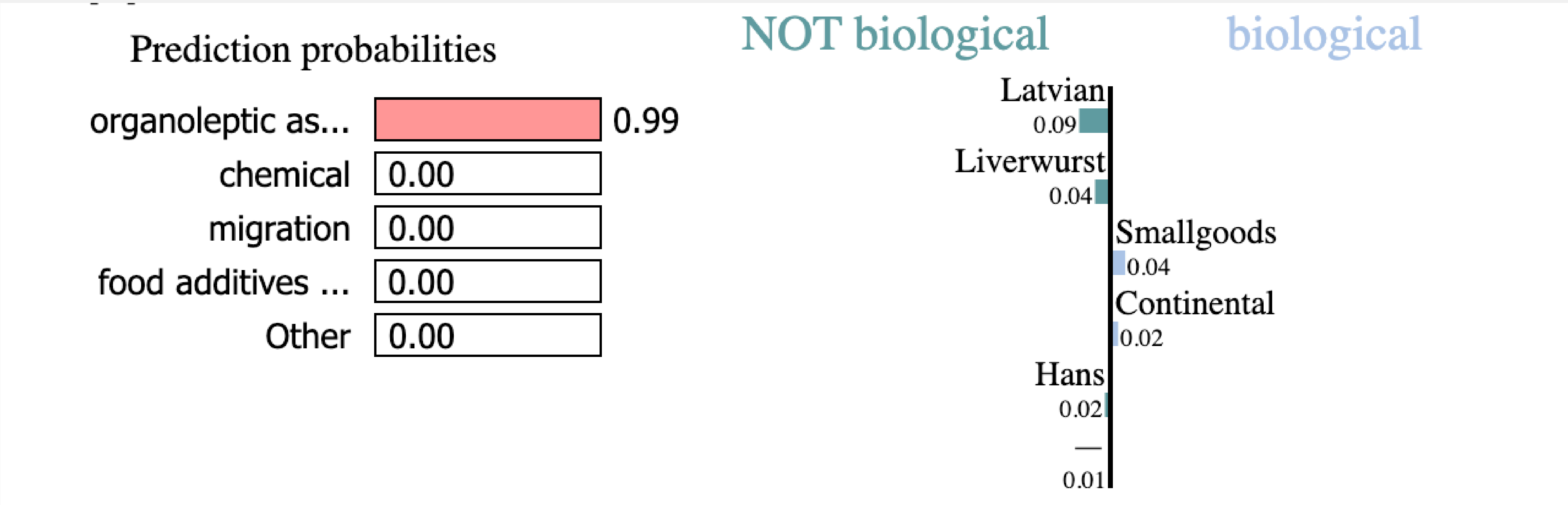
Researched Approaches

- Explainability Models
- LIME/SPLIME
- SHAP

Issues faced with Researched Approach

- Computationally very expensive
- Difficulty in finding implicit products/hazards.
- Complex in Implementing

LIME



Text with highlighted words

Hans Continental Smallgoods — Latvian Liverwurst

```
INFO:root:Example 3/5051 start
['Recall', 'Notification', 'FSIS', '-011-98']
Recall Notification: FSIS-011-98
['Recall', 'Notification', ':', 'FSIS', '-011-98']
```

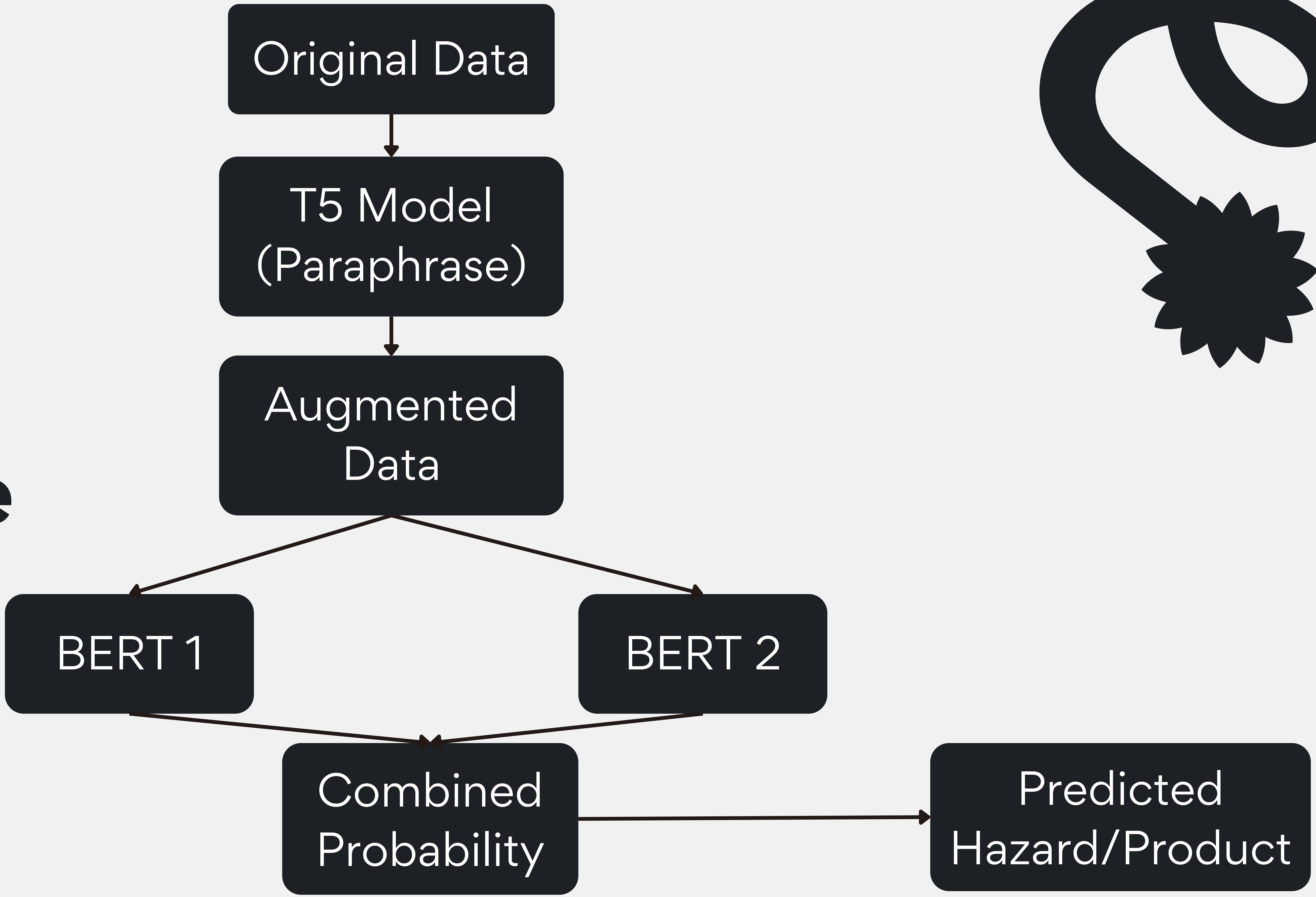
Our Approach

- Augment upto a threshold for each class
- T5 model for paraphrasing
- Classification using BERT
- Ensembling

Augmentation Sizes

- 2,120 Rows for Hazard
- 5,111 Rows for Product

Task Pipeline



Baseline BERT Outputs - Hazard

	precision	recall	f1-score	support
Aflatoxin	0.00	0.00	0.00	4
alcohol content	0.00	0.00	0.00	1
alkaloids	0.00	0.00	0.00	2
allergens	0.00	0.00	0.00	4
almond	1.00	0.07	0.13	14
undeclared additive	0.00	0.00	0.00	2
undeclared constituent	0.00	0.00	0.00	1
virus	0.00	0.00	0.00	1
walnut	0.40	0.40	0.40	5
accuracy			0.52	1017
macro avg	0.10	0.11	0.09	1017
weighted avg	0.45	0.52	0.43	1017

Enhanced Local Outputs - Hazard

	precision	recall	f1-score	support
Aflatoxin	1.00	1.00	1.00	3
abnormal smell	1.00	0.67	0.80	3
alcohol content	1.00	1.00	1.00	3
alkaloids	1.00	1.00	1.00	3
allergens	0.67	0.67	0.67	3
almond	1.00	1.00	1.00	7
altered organoleptic characteristics	1.00	1.00	1.00	3
amygdalin	1.00	0.67	0.80	3
antibiotics, vet drugs	1.00	1.00	1.00	3
bacillus spp.	1.00	0.33	0.50	3
bad smell / off odor	1.00	1.00	1.00	3
bone fragment	1.00	0.67	0.80	3
brazil nut	1.00	0.67	0.80	3
bulging packaging	0.60	1.00	0.75	3
cashew	1.00	1.00	1.00	3
celery and products thereof	1.00	1.00	1.00	3
cereals	1.00	1.00	1.00	3
cereals containing gluten and products thereof	0.91	0.95	0.93	21

accuracy			0.90	721
macro avg	0.86	0.85	0.84	721
weighted avg	0.90	0.90	0.89	721

Baseline BERT Outputs - Product

	precision	recall	f1-score	support
Catfishes (freshwater)	0.00	0.00	0.00	5
Fishes not identified	0.00	0.00	0.00	6
Not classified pork meat	0.00	0.00	0.00	3
whisky	0.00	0.00	0.00	1
white lasagna sauce	0.00	0.00	0.00	1
wine	0.00	0.00	0.00	2
wraps	0.00	0.00	0.00	4
yoghurt	0.00	0.00	0.00	5
accuracy			0.09	1017
macro avg	0.00	0.01	0.00	1017
weighted avg	0.02	0.09	0.03	1017

Enhanced Local Outputs - Product

	precision	recall	f1-score	support
Catfishes (freshwater)	1.00	1.00	1.00	1
Dried pork meat	1.00	1.00	1.00	1
Fishes not identified	0.67	1.00	0.80	4
Groupers (generic)	1.00	1.00	1.00	1
Not classified pork meat	1.00	1.00	1.00	2
white mustard seeds	1.00	1.00	1.00	1
whole chicken	1.00	1.00	1.00	1
wine	1.00	1.00	1.00	2
wraps	1.00	1.00	1.00	2
yellow peas	1.00	1.00	1.00	1
yoghurt	0.75	1.00	0.86	3
yogurt raisins	0.00	0.00	0.00	1
accuracy			0.83	1232
macro avg	0.77	0.81	0.78	1232
weighted avg	0.77	0.83	0.79	1232

SemEval Score - ST2

#	User	Entries	Date of Last Entry	Team Name	Score Sub-Task 1 ▲	Score Sub-Task 2 ▲	Detailed Results
1	pinganxiaofendui	19	11/26/24	Pingan AI Team	0.0000 (1)	0.5118 (1)	View
2	polarisshi	16	11/27/24	PingAn AI - NLP	0.0000 (1)	0.4791 (2)	View
3	jerryk42	3	11/26/24		0.0000 (1)	0.4515 (3)	View
4	meesuma5	5	11/27/24		0.0000 (1)	0.4510 (4)	View

Future Directions

- Try to map hazard/product-categories to hazard/product and limit possibilities.
- Synthesize more data for both hazards and products.

References

- [1] Marco Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pages 97–101, San Diego, California. Association for Computational Linguistics.
- [2] Buyuktepe, O., Catal, C., Kar, G., Bouzembrak, Y., Marvin, H., & Gavai, A. (2023). Food fraud detection using explainable artificial intelligence. Expert Systems, e13387.
- [3] Ozyegen, O., Jahanshahi, H., Cevik, M. et al. Classifying multi-level product categories using dynamic masking and transformer models. J. of Data, Inf. and Manag. 4, 71–85 (2022).
- [4] Yu Meng, Yunyi Zhang, Jiaxin Huang, Chenyan Xiong, Heng Ji, Chao Zhang, and Jiawei Han. 2020. Text Classification Using Label Names Only: A Language Model Self-Training Approach. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Online. Association for Computational Linguistics.
- [5] John Pavlopoulos, Leo Laugier, Alexandros Xenos, Jeffrey Sorensen, and Ion Androutsopoulos. 2022. From the Detection of Toxic Spans in Online Discussions to the Analysis of Toxic-to-Civil Transfer. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 3721–3734, Dublin, Ireland. Association for Computational Linguistics.
- [6] Assael, Y., Sommerschield, T., Shillingford, B. et al. Restoring and attributing ancient texts using deep neural networks. Nature 603, 280–283 (2022).

Thank You

Q & A