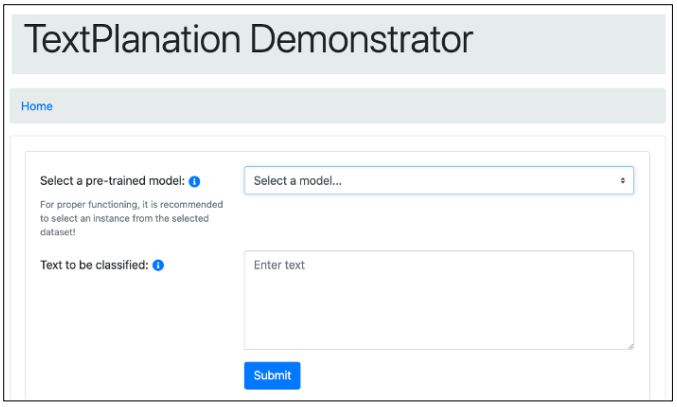
# TextPlanation: Explainable Machine Learning for Text Data

## MARKET NEED

- Natural language Processing (NLP) combined with machine learning has been used for many applications such as *sentiment analysis*, *topic modelling, document categorisation, spam detection, fake news identification,* and *chatbots*.
- Explaining why a machine learning model made a particular decision is key to trusting AI based systems as well as complying with legislation such as the GDPR.
- Although a number of different explainable machine learning libraries and tools have been developed, it is often unclear as to which to use in different scenarios, and how these work specifically in the context of text data.



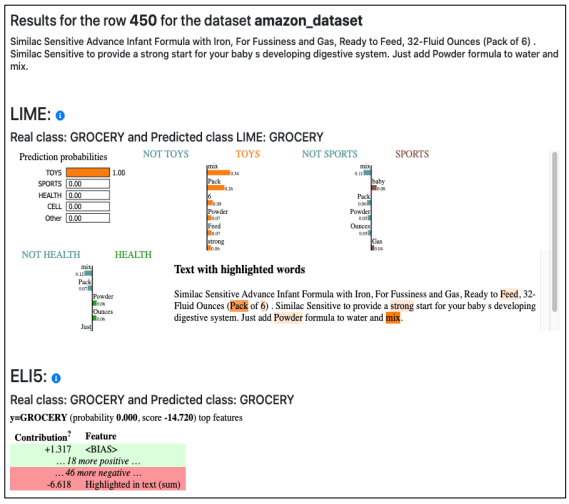

## TECHNOLOGY SOLUTION

- The TextPlanation demonstrator displays the visualised output of five different explainable ML libraries: LIME, SHAP, LRP, SKATER, and ELI5.
- These explain the decisions made by text classifiers trained over a number of sample datasets.
- The tool can be used to understand the benefits and limitations of these open-source libraries in different contexts and with different types of text data.

## KEY FEATURES

The key features of TextPlanation are:

- The user can choose from a range of pre-trained text classification models.
- The user can enter their own text to be classified, and the explanations of the model decision are presented.
- Outputs from each of five explainable ML libraries are shown on the same screen allowing direct comparison.
- New explainable ML libraries can easily be integrated.




## CONCLUSION

CeADAR's TextPlanation demonstrator shows how five different open-source explainable ML libraries explain text classifier model decisions. This can be used to understand the benefits and limitations of these libraries in different contexts and with different types of text data


**RESEARCH TEAM**

CeADAR Applied Research Group, UCD


**www.ceadar.ie**