Department	School of Computing
Supervisors	Dr Pierre Le Bras
Project Title	Trustworthy Exploration of Large Portfolio with Ensemble Topic Mapping

PROJECT DESCRIPTION

This project aims to explore the affordances and limitations of an ensemble approach to topic modelling and the opportunities and challenges in brings to the visualisation of such models for non-expert end-users. While powerful in their ability to construct complex thematic models of large text portfolio, probabilistic topic modelling methods are stochastic by nature. This trait often leads to instability in the computed model. This issue also often lead to several interpretability and explainability challenges faced by non-expert end-users.

This project primary objective is to create a learning framework that would take on an ensemble approach to topic modelling. While the aim of such approach primarily focuses on stabilising the topic discovery, the project would also seek to use it for improving metrics such as coherence and perplexity. Research and evaluation would be performed to study suitable output fusion methodologies.

From this, the project will investigate and compare the effects of an ensemble learning approach to the further data analysis processes that stem from topic modelling, for example, topic similarity, document relevance ranking, or trend analysis. A topic stability metric greatly impact the traditional approaches to such analyses, typically by creating an importance ranking of topics.

In the end the project should the see the development of an intuitive and interactive Topic Mapping viualisation dashboard that allows non-expert end-users to explore their portfolio and gain greater insights from the ensemble model. The final objective of this project will be to evaluate the interpretability and explainability of these stochastic ensemble models and their visualisations.

Prospective applicants are encouraged to contact the Supervisor before submitting their applications. Applications should make it clear the project you are applying for and the name of the supervisor(s).

Academic qualifications

A first degree (at least a 2.1) ideally in Computer Science or related areas, with a good fundamental knowledge of Data Science or Machine Learning.

English language requirement

IELTS score must be at least 6.5 (with not less than 6.0 in each of the four components). Other, equivalent qualifications will be accepted. Full details of the University's policy are available online.

Essential attributes:

- Experience of fundamental concepts in Data Analytics and Visualisation
- Competent in programming (e.g., Python/Java) with a good knowledge of web languages (e.g., JavaScript)
- Knowledge of Machine Learning methods and user studies
- Good written and oral communication skills
- Strong motivation, with evidence of independent research skills relevant to the project
- Good time management

Desirable attributes:

Experience with building data analysis pipeline and/or developing interactive visualisation dashboards for the web (e.g., D3.js) is a plus.

Indicative Bibliography	Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84. Chuang, J., Roberts, M. E., Stewart, B. M., Weiss, R., Tingley, D., Grimmer, J., & Heer, J. (2015). TopicCheck: Interactive alignment for assessing topic model stability. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 175-184). Le Bras, P., Robb, D. A., Methven, T. S., Padilla, S., & Chantler, M. J. (2018, April). Improving user confidence in concept maps: Exploring data driven explanations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
Enquiries	For informal enquiries about this PhD project, please contact Dr Pierre Le Bras – P.LeBras@napier.ac.uk
Web page	https://www.napier.ac.uk/research-and-innovation/research-degrees/application-process