

Company Name:	Experian
Team / Department:	DataLabs
Address:	Kings Cross, London

Provisional title for project:

Transductive Learning for Text Classification

Short abstract of what the project would probably entail, and the data to be used:

Transductive learning is the process of using an unlabeled test set for improving the performance of a machine learning algorithm. This project proposes to study transductive learning algorithms (e.g. in “Reliable Classifications with Machine Learning”, Kurak and Kononenko, 2002) on the Twitter Sentiment140 dataset and the News Aggregator dataset. To generate a certain dataset split, the tweets will be grouped by date and user and news articles will be grouped by publisher. The student will first conduct a literature review, and then explore non-transductive approaches (TFIDF/Naïve Bayes/Random forests) as well as approaches to transductive learning, starting with simple variations in sampling strategies. The results are anticipated to be useful to other text classification problems in test data is generated from a source different to the training, whilst salient features are constant.

Essential and desirable skills that the student would need to have:

Essential skills (NB): Good knowledge Python programming and familiarity with Machine Learning

Desirable: Knowledge of text classification.

Preferred degree programmes (if any): Computer science, mathematics

Would any work by the student need to be carried out on site at the Company with the exception of supervisory meetings?

No

Any issues of data confidentiality and IPR that would need to be resolved?

No

Preferred selection method

F2f Interview

Support and training offered by the company

Regular weekly or biweekly meetings with the team in their London office.

Financial assistance offered by the company

I agree to pay the student £500 (plus travel expenses) ...

For details on how to apply, please visit:

<https://www.cdrc.ac.uk/retail-masters/details-for-students/>