Proposer: Ted Scully	**Contact:** ted.scully@mtu.ie	Student Contact: OK

Project Title:

Identification of anonymous authors using textual analysis and machine learning

Project Abstract:

Stylometry is a text categorization problem that attempts to identify the author or profile (age, gender, etc) of anonymous writings. The written text could be anything from social media tweets and blog posts to entire books. There are three main categories of problems that are addressed in this area [1]:

1) Authorship identification (authorship attribution): It determines the likelihood of a piece of writing to be produced by a particular author by examining other writings by that author.
2) Authorship characterization: It summarizes the characteristics of an author and generates the author profile based on his/her writings like age, gender, educational.
3) Similarity detection: It compares multiple pieces of writing and determines whether they were produced by a single author without actually identifying the author like Plagiarism detection

The objective of this project is to investigate the application of textual analysis techniques coupled with machine learning algorithms such as SVMs, Clustering Analysis and Bayesian techniques to one three problems listed above.

1. S. Nirkhi, R. Dharaskar. Comparative study of Authorship Identification Techniques for Cyber Forensics Analysis. International Journal of Advanced Computer Science and Applications 2013

Technologies Utilised:
Python, Scikit-Learn