Hamilton Institute Student Seminar Series

**Navoda Senavirathne, University of Skövde, Sweden**

**Title:** On the challenges of privacy preserving machine learning in the context of data anonymization

**Abstract:** The requirement for privacy preserving machine learning (PPML) has surfaced with the usage of sensitive data to train the machine learning models, and with the emergence of different threats exploiting the privacy vulnerabilities in machine learning. Most of the prominent PPML techniques such as differential privacy operates under the assumption that access to original data are available for training the machine learning models. However, given the increasing threats towards privacy, it is plausible that the data owners attempt to anonymize the data before they are being released for analysis to avoid any privacy violations. This process is known as the Privacy Preserving Data Publication (PPDP). Anonymization is used to achieve PPDP which introduces a controlled distortion to data. This degrades the data utility,  and subsequently, the quality of any analysis done on anonymized data will also be adversely affected. Therefore, understanding the impact of data anonymization on machine learning is important. Nevertheless, it is paramount to consider whether the models trained on anonymized data can mitigate the privacy vulnerabilities faced by machine learning successfully. This talk will mainly focus on the interplay between PPDP and PPML in terms of privacy-utility trade-off and the existing challenges in this context.