



جامعة الإمارات العربية المتحدة  
United Arab Emirates University

The College of Graduate Studies and the College of Information  
Technology Cordially Invite You to a

**PhD Dissertation Defense**

Entitled

***Streaming Feature Grouping and Selection (SFGS) for Classification Problem***

by

**Noura Helal Hamad Khudouma Al Nuaimi**

Faculty Advisor

**Dr. Mohammad Mehedy Masud, Department of Information Systems and Security  
College of Information Technology**

Date & Venue

**11:00 AM**

**Thursday, 28 March 2019**

**Room 1036, E1 Building**

**Abstract:** Real-time data has always been an important element for organizations. The quickness with which data is delivered is critical to their businesses. Today organizations understand the importance of real-time data analysis to maintain benefits from their generated data. Real-time data analysis is also known as real-time analytics, streaming analytics, real-time streaming analytics, and event processing. Stream processing is the key to have results in real time. Stream processing allows us to process the data stream in real time as they arrive. The concept of streaming data means the data are generated dynamically, where the full stream is unknown or even infinite. In machine learning, streaming feature selection has always been a preferable method in the preprocessing of streaming data. Recently, feature grouping that could measure the hidden information between selected features has begun gaining attention. The main contribution of this dissertation is in solving the issue of the extremely high dimensionality of streaming data by delivering a streaming feature grouping and selection algorithm. The literature review presents a comprehensive review of the current streaming feature selection approaches and highlights the state-of-the-art algorithms trending in this area. In addition, the literature review contrasts a detailed comparison of these approaches. Our proposed algorithm is designed with the idea of grouping similar features to reduce redundancy and to handle the stream of features in an online fashion. We have implemented this algorithm and evaluated it using benchmark datasets against state-of-the-art streaming feature selection algorithms and feature grouping techniques. The results showed better

performance regarding prediction accuracy comparing with state-of-arts algorithms.

**Keywords:** Stream of features, features grouping, feature selection, relevance analysis, redundancy analysis.