

3rd International Conference on

Big Data Analysis & Data Mining

September 26-27, 2016 London, UK

Application of feature selection algorithms for life sciences

Geervani Koneti

Tata Consultancy Services Ltd, India

Pharmaceutical industry is faced with increased attrition of drug candidates at late stages of drug development mainly due to poor drug safety and efficacy profiles. Thus, there is growing interest within pharmaceutical industries to explore and exploit new innovative approaches to reduce the attrition rates and consequently, the overall cost of discovery of a drug. One such approach that can now, compared to the past, be realized more effectively is integrated and big data analytics. In an effort to address the R&D productivity, we are in process of developing a data analytics platform to facilitate predictive model building, informed decision making using analytics and effortless transition across various stages of drug discovery and development. While the exponential increase in data presents its own challenges, we are enhancing our models to utilize the advanced computing technology to perform the required analytics in real time. One such example we will discuss is parallelization of exhaustive search for feature selection and modelling. In the absence of the advanced computing resources, the frequently employed alternative is the use of heuristic methods, which often trade accuracy for time. However, it is significant to note that the reproducibility of most of these methods is questionable and few pharmaceutical companies are devising strategies to counter this issue too. Another approach we will be discussing in the talk is how to improve the reproducibility and reliability of the modelling techniques, in specific the random based feature selection methods used for model building.

k.geervani@tcs.com