

**PREDICTING THE MOST EFFECTIVE SRI LANKAN
BOWLER FOR THE TARGETED BATSMAN IN CRICKET
USING MACHINE LEARNING**

W P G G SANJU DE SILVA

A dissertation submitted in partial fulfillment for the requirement for Bachelor of Science
(Honours) degree in Software Engineering

Department of Computing

**Informatics Institute of Technology, Sri Lanka in Collaboration with
University of Westminster, UK**

2020

Abstract

This research presents findings of a study to predict the most effective bowler to a targeted batsman. We use Naive Bayes (NB) approach to make this prediction using the data collected with 15 features, comprised of variables related to batting, bowling, team composition, venue of the match and other. Upon the construction of an initial model, our objective is to improve the accuracy of predicting the winner using some feature selection algorithms, namely univariate, recursive elimination, and principle component analysis (PCA). Furthermore, we examine the contribution of the appropriate ratios of training sample size to testing sample size on the accuracy of prediction. According to the experimental findings, the accuracy of this prediction system can be improved with the use of feature selection algorithm. Moreover, the accuracy of winner prediction becomes the highest (65.71%) with the univariate feature selection method, compared to its counterparts. By selecting the appropriate ratio of the sample sizes of training sample to testing sample, the prediction accuracy can be further increased.