

## **[Keynote Speech]: Feature Selection and Predictive Modeling of Housing Data Using Random Forest**

**Authors :** Bharatendra Rai

**Abstract :** Predictive data analysis and modeling involving machine learning techniques become challenging in presence of too many explanatory variables or features. Presence of too many features in machine learning is known to not only cause algorithms to slow down, but they can also lead to decrease in model prediction accuracy. This study involves housing dataset with 79 quantitative and qualitative features that describe various aspects people consider while buying a new house. Boruta algorithm that supports feature selection using a wrapper approach build around random forest is used in this study. This feature selection process leads to 49 confirmed features which are then used for developing predictive random forest models. The study also explores five different data partitioning ratios and their impact on model accuracy are captured using coefficient of determination (r-square) and root mean square error (rsme).

**Keywords :** housing data, feature selection, random forest, Boruta algorithm, root mean square error

**Conference Title :** ICDSA 2017 : International Conference on Decision Sciences and Analytics

**Conference Location :** Boston, United States

**Conference Dates :** April 24-25, 2017