

**The Power of Random: Using Perturbation Experiments to Improve  
Model Accuracy and Interpretation  
November 7, 2018 – AT&T Conference Center, Austin TX**

**PRELIMINARY AGENDA – Subject to Change**

<b>Registration Coffee &amp; Snacks</b>	<b>09:00 – 10:00</b>	
Randomization Principles for Sampling	10:00 – 10:30	Simple Random Sampling; Stratified Sampling; Bootstrap Sampling (Cross-validation)
Data Sets Overview	10:30 – 10:55	Describe the two data sets: cup98 and Titanic. Walk through data definitions and load data into KNIME Analytics Platform Create Statistics Nodes to summarize the data
Randomization Principles for Sampling	10:55 – 11:15	Simple Random Sampling; Stratified Sampling; Bootstrap Sampling (Cross-validation)
Randomization for Missing Value Imputation	11:15 – 12:00	Missing Values; Mean Imputation (Missing Value Node). Random imputation; Random Referential Imputation
Randomization in Building Models (Bagging, Random Forests)	12:00 – 13:00	Ensembles; Bagging (Bootstrap Sampling); Bagged Logistic Regression (CV Ensembles?); Random Forests (Random Variable Selection)
<b>Lunch</b>	<b>13:00 – 14:00</b>	
Randomization to Assess Confidence in Model Accuracy	14:00 – 14:35	Target Shuffling
Randomization to Assess Variable Importance	14:35 – 15:30	Input Shuffling
<b>Break</b>	<b>15:30 – 16:00</b>	
Randomization to Create Model Prediction Confidence Intervals	16:00 – 17:00	CI Background Metrics such as Input Variable Selection; Logistic Regression Coefficients (as Baseline)
Final Q&A	17:00 – 18:00	

