Slide 1: Title Slide

"Welcome to our presentation, 'Predicting Bullying Victims in Schools.' This study, conducted by SiCheng Yi and Amina Bauyrzhan, utilizes the 2013 National Crime Victimization Survey for insightful data analysis."

Slide 2: Project Outline

"Our objective is to predict bullying victims using the 2013 school crime supplement dataset. Our methodology includes data preprocessing, feature selection, and model evaluation, using R packages like ROSE, caret, and e1071."

Slide 3: About the Dataset

"The dataset from the 2013 School Crime Supplement provides a comprehensive view of school safety, crime perception, and bullying experiences. It's a valuable resource for understanding the dynamics of bullying in school settings."

Slide 4: Data Preprocessing and Feature Selection

"In data preprocessing, we handled missing values, normalized features, and converted categorical data into numerical formats. Using Random Forest, we identified the top 20 features for bullying prediction, which helped in focusing on the most influential variables."

Slide 5: Feature Selection with Random Forest

"The Random Forest algorithm was pivotal in feature selection, given its robustness for high-dimensional data. We used %IncMSE and IncNodePurity to determine feature importance, streamlining our predictive model."

Slide 6: Classification Models

"We analyzed several classification models: Logistic Regression, Decision Tree, Random Forest, SVM, k-NN, and Naive Bayes. Each model was chosen for its unique properties, here is the detail of the models.

Slide 7: Model Evaluation Metrics

"To evaluate our models, we used metrics like True Positive Rate, False Positive Rate, Precision, F-measure, and ROC Area. These metrics were crucial in ensuring the accuracy and reliability of our predictions."