**PREDICTION OF INDIVIDUAL INCOME**

**USING A MACHINE LEARNING MODEL**

**G SAI LAHARI**

**SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY**

**Abstract Report**

The main objective of the project is to create a classification model to predict whether a person makes over $50k a year. The data was extracted from the 1994 Census bureau database by Ronny Kohavi and Barry Becker (Data mining and visualization, Silicon Graphics). The data sample of shape (32561,15).The data is taken and handled null values in any of the columns. Then, the data is split into training and testing sets with 20% as testing size for every model. The classification prediction model is created for Decision Tree, Random Forest Classifier, Logistic Regression, KNN Classifier, and SVC Classifier with linear kernel. The misclassification is also derived for every model through computing confusion matrix.

*Misclassification = (FP+FN) / (Sample Size)*

The classification report is also generated for each of these models. After creating different models, it is found that the model with best accuracy was Random Forest Classifier with 86% accuracy.