Homework 7 - Classification

Using the provided Census Income Dataset, create an ML pipeline in spark to predict whether or not an adult will have an income greater than 50K. Information about the dataset can be found at: https://archive.ics.uci.edu/dataset/2/adult

For complete credit your model pipeline should have the following:

1. Feature transformations
2. Train/test split
3. Any classifier that predicts whether or not the adult will have an income below or above 50k.
4. Evaluation metrics such as confusion matrix and F1 score showing your model’s accuracy on the train and test sets.