

ML Case Study

Please use Python for the following. Document your steps and state any assumptions you make. A single Jupyter notebook file (.ipynb) with your approach and code is sufficient. Make sure it is commented and easy to follow. If you prefer, you can submit a text document with your analysis too.

The data for the case study comes from the **Census Income data set** in the UCI Machine Learning Repository. **Your objective is to predict whether income of an individual exceeds \$50K/per year based on census data.** More importantly the objective for you is to demonstrate how you tackle a traditional supervised learning problem.

- 1) Download the data from the following link <http://archive.ics.uci.edu/ml/datasets/Adult> into Python.
- 2) Read the Attribute Information from the web page. More information can be found in the Data Set Description.
- 3) Be sure to document your steps and your rationale behind each one of them. This is best done with Jupyter notebook or a word document.
- 4) Predict the class variable using the given attributes. Use two supervised learning methods. Avoid over-fitting.
- 5) Assess the performance of your models using appropriate metrics. Discuss which model is best and why.
- 6) Please provide an explanation in plain English of the business problem you were trying to solve and your analysis of the outcomes. Please add any supporting visualizations to support your explanation.
- 7) This is intentionally left somewhat open-ended. We want to see how you approach the problem.

Please Note: We expect you to approach this business problem with honesty and integrity and rely on your knowledge of machine learning. Please do not copy from any web-based repository. Candidates will be disqualified if we notice attempts of plagiarism.