Adult Census Income Analysis
Team 6

Lecturer: Dr. Arash Azarfar

Team members:

Dina Omidvar Tehrani, Niloofar Tavakolian

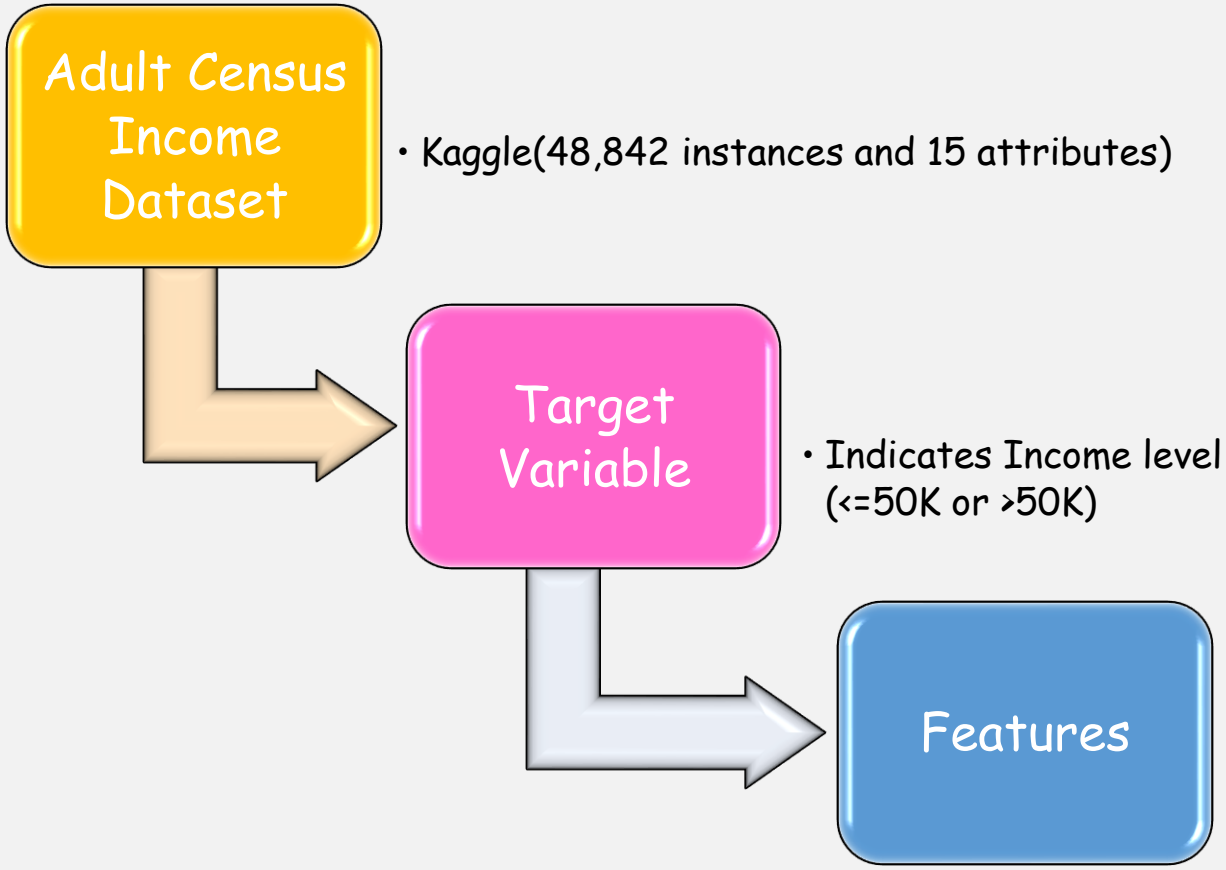
Nastaran Naseri, SeyedehMojdeh Haghighat Hosseini

Overview of the problem

Have you ever wondered what key factors could propel your income to the next level?

- Income Patterns: Dissecting Adult Income Factors
- Income Prediction: Lighting Up Inequality & Empowerment
- Demographic Insights: Probing Socioeconomic Dynamics

Dataset Selection



Expectations and Goals

Developing a predictive model

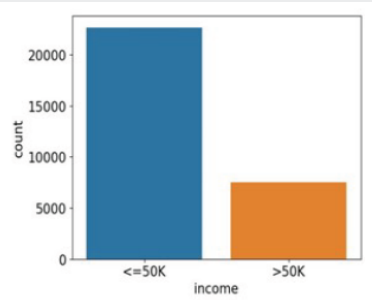
Evaluating model performance

Comparing methodologies

Feature importance analysis

Generating actionable insights

Pre-processing

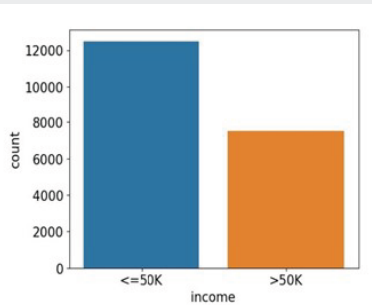


(a) Imbalance dataset

Dealing with class imbalance

Handling missing values

Dealing with categorical features



(b) Balanced dataset

Supervised Learning: Decision Tree classification

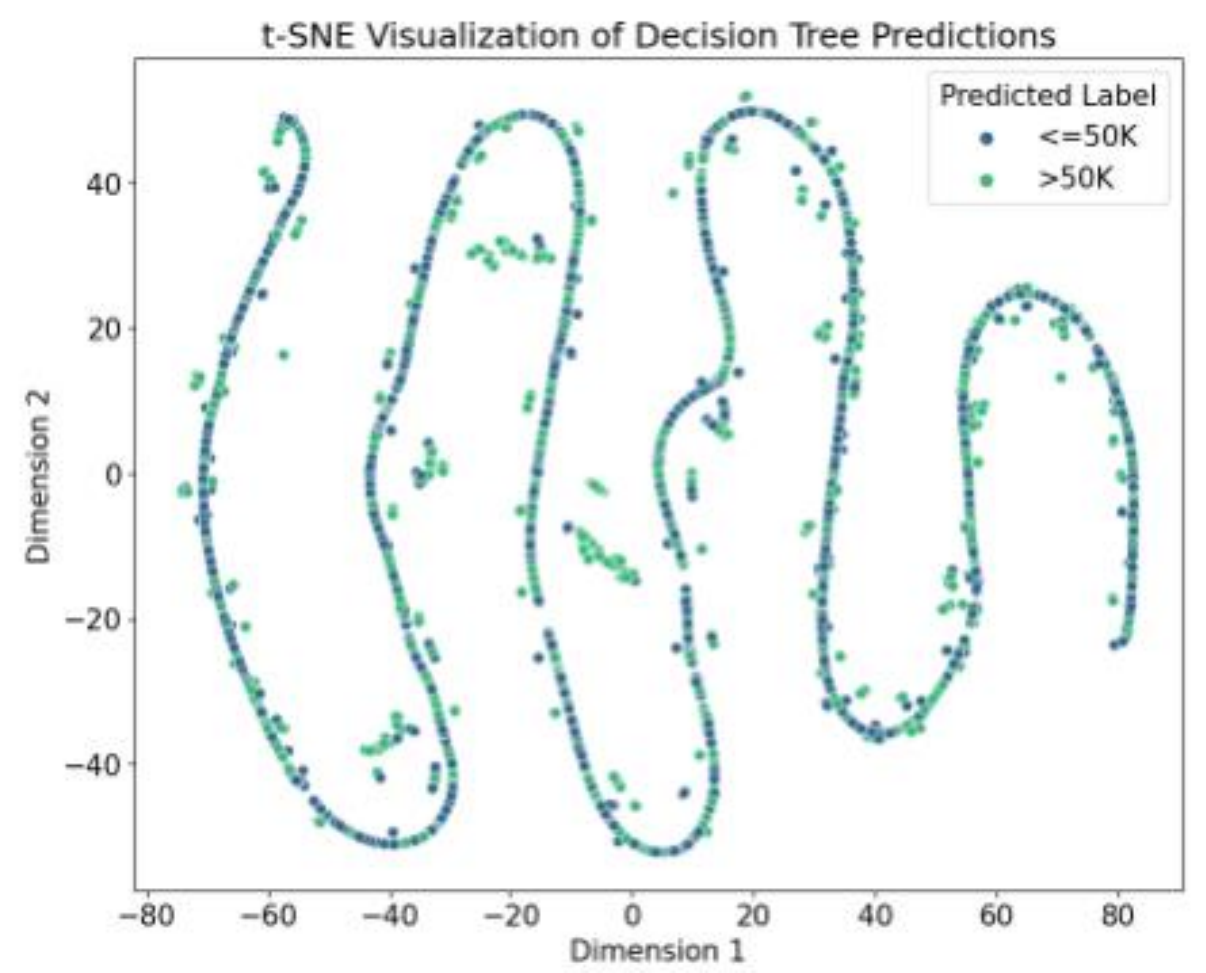
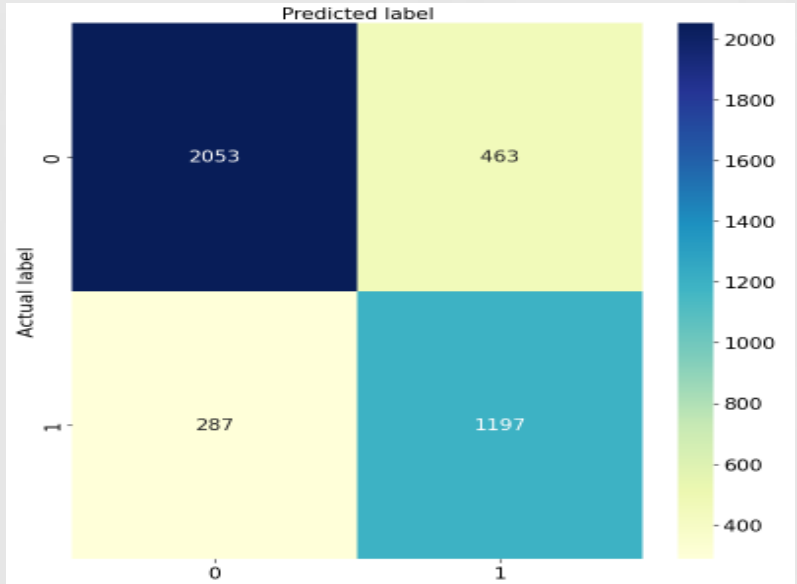


Table 2: Evaluation of Model's Performance (Supervised learning Classification with Decision Trees) [8]

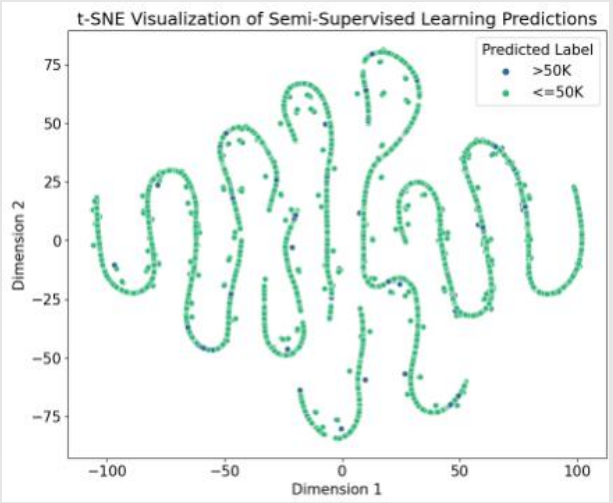
	Precision	Recall	F1-score	Support
≤50K	0.88	0.82	0.85	2516
>50K	0.72	0.81	0.76	1484
Accuracy			0.81	4000
Macro Avg	0.80	0.81	0.80	4000
Weighted Avg	0.82	0.81	0.81	4000

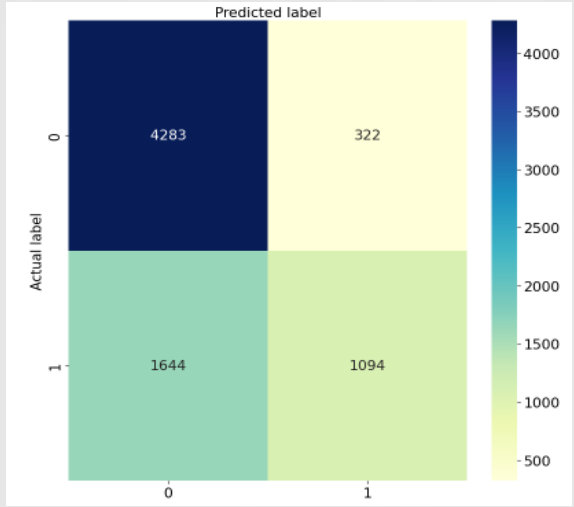


Semi-supervised Learning Using a Decision Tree Classifier

Table 4: Evaluation of Model's Performance (Semi-supervised learning Classification with Decision Trees) [8]

	Precision	Recall	F1-score	Support
≤50K	0.72	0.93	0.81	4605
>50K	0.77	0.40	0.53	2738
Accuracy			0.73	7343
Macro Avg	0.75	0.66	0.67	7343
Weighted Avg	0.74	0.73	0.71	7343



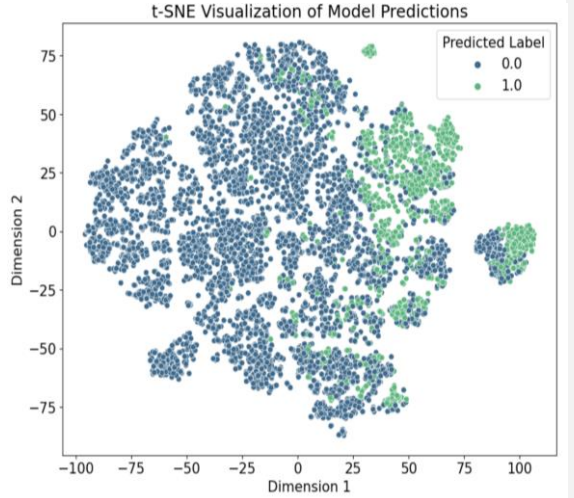


Supervised Learning: Classification with a Deep Learning Model

	Layers		
Fully connected	FC1,ReLU	Featues in:34	Features out: 64
Dropout	DO1		
Fully connected	FC2. ReLU	Featues in:64	Features out: 128
Dropout	DO2		
Fully connected	FC3. ReLU	Featues in:128	Features out: 64
Dropout	DO3		
	FC4. BCE	Featues in:64	Features out: 1

Table 6: Evaluation of Model's Performance (DNN) [8]

	Precision	Recall	F1-score	Support
≤50K	0.8845	0.9345	0.9088	7415
>50K	0.7488	0.6155	0.6757	2354
Accuracy			0.8576	9769
Macro Avg	0.8167	0.7750	0.7922	9769
Weighted Avg	0.8518	0.8576	0.8526	9769



Conclusion

Predicting Income

Data pre-processing:

Handling missing values, imbalanced dataset, and categorical features

Decision Tree

Accuracy: 81.1

Semi-Supervised Learning

Accuracy:73.24

DNN Model

Accuracy:85.76

Refrences

- “Adult census income,” kaggle, https://www.kaggle.com/datasets/uciml/adult- census-income?resource=download.
- “Handling missing data,” scikit-learn documentation, 2023, https://scikit-learn.org/stable/modules/impute.html.
- D. Omidvar, N. Tavakolian, N. Naseri, and S. H. Hosseini, “Project github,” https://github.com/dinaomidvartehrani/Applied- AI- .git.
- “Resampling methods,” scikit-learn documentation, 2023, https://scikit-learn.org/stable/modules/classes.html#module- sklearn.utils.
- J. Brownlee, “Why one-hot encode data in machine learning?” 2020, https://machinelearningmastery.com/why-one-hot-encode-data-in-machine- learning/.