

Project1. Income prediction

Dataset description

Attribute Name	Definition	Example
age	Age (years)	38, 42, 71
workclass	Workclass 8 different categories: (Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked)	"Private", "Local-gov", "Never-worked"
fnlwgt	Final Weight*	83311, 338409
education	Education: (Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool)	"Bachelors", "9th", "Preschool"
education-num	Years of education	13, 9, 7
marital-status	Marital Status: (Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse)	"Divorced", "Separated", "Widowed"
occupation	Occupation: (Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces)	"Tech-support", "Armed Forces", "Sales"
relationship	Relationship: (Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried)	"Wife", "Unmarried", "Own-child"

race	Race: (White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black)	"White", "Asian-Pac-Islander", "Other"
sex	Sex: (Male, Female)	Male, Female
capital-gain	Amount of capital gained	14084, 0, 5178
capital-loss	Amount of capital lost	0, 2042, 1902
hours-per-week	Number of hours worked per week	40, 50, 70
native-country	Native country: (United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinidad&Tobago, Peru, Hong, Holand-Netherlands)	"China", "Italy", "Vietnam"
Income (Target column)	Either the income is greater than \$50,000 or lesser than and equal to \$50,000: (>50K, <=50K)	">50K", "<=50K"

Project Requirements.

1. Preprocessing: Before building your classification models, you need to make sure that the dataset is processed, clean and ready-to-use.
2. Feature selection: Try different feature selection techniques.
3. Classification:
 - Train at least 3 different classification models (e.g., Logistic regression, SVM and Decision tree, etc).
 - Choose at least two hyperparameters to vary, and explain the impact of varying those two hyperparameters with different values.
 - Train the models on the given training set and evaluate the model on the given testing set.

4. Model Evaluation: You should try different evaluation metrics to evaluate your model.

- Accuracy
- Precision & recall
- F1-score
- Confusion matrix