Prediction Of Adult Income based on Census Data

###### A PROJECT REPORT

*Submitted by*

Sameer Singh [RA2111003010750]

Ansh Singh [RA2111003010749]

Kshitij Gautam [RA2111003010755]

Aditya Raj [RA2111003010777]

*Under the Guidance of*

Dr. Godfrey Winster S

Associate Professor, Department of Computing Technologies

*in partial fulfillment of the requirements for the degree of*

#### BACHELOR OF TECHNOLOGY

in

#### COMPUTER SCIENCE ENGINEERING

with specialization in Computer Science and Engineering



#### DEPARTMENT OF COMPUTATIONAL INTELLIGENCE COLLEGE OF ENGINEERING AND TECHNOLOGY SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY

KATTANKULATHUR- 603 203

###### OCTOBER 2023

Annexure II

Department of Computational Intelligence

SRM Institute of Science & Technology Own Work\* Declaration Form

To be completed by the student for all assessments

**Degree/ Course :** 18CSE355T - Data Mining and Analytics

**Student Name :** SAMEER SINGH

**Registration Number :** RA2111003010750

**Student Name :** ANSH SINGH

**Registration Number :** RA2111003010749 **Student Name :** KSHITIJ GAUTAM

**Registration Number :** RA2111003010755

**Student Name :** ADITYA RAJ

**Registration Number :** RA2111003010777

**Title of Work :** Prediction of Adult Income based on Census Data

I / We hereby certify that this assessment compiles with the University’s Rules and Regulations relating to Academic misconduct and plagiarism\*\*, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / We have met the following conditions:

* Clearly referenced / listed all sources as appropriate
* Referenced and put in inverted commas all quoted text (from books, web, etc.)
* Given the sources of all pictures, data etc. that are not my own
* Not made any use of the report(s) or essay(s) of any other student(s) either past or present
* Acknowledged in appropriate places any help that I have received from others (e.g., fellow students, technicians, statisticians, external sources)
* Compiled with any other plagiarism criteria specified in the Course handbook / University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

|  |
| --- |
| **DECLARATION:** |
| I am aware of and understand the University’s policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except were indicated by referring, and that I have followed the good academic practices noted above.  Sameer Singh [RA2111003010750]  Ansh Singh [RA2111003010749]  Aditya Raj [RA2111003010777]  Kshitij [RA2111003010755] |
| If you are working in a group, please write your registration numbers and sign with the date for every student in your group. |



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR – 603 203

#### BONAFIDE CERTIFICATE

Certified that 18CSE355T project report titled “**Prediction of Adult Income based on Census Data**” is the bonafide work of “Sameer Singh [RA2111003010750], Ansh Singh [RA2111003010749], Kshitij Gautam [RA2111003010755], Aditya Raj [RA2111003010777]”

who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE SIGNATURE**

**Faculty In – Charge**

Dr. Godfrey Winster S Associate Professor

Department of Computing Technologies SRM Institute of Science & Technology

**HEAD OF THE DEPARTMENT**

Dr. M. Pushpalatha Professor & Head

Department of Computing Technologies SRM Institute of Science & Technology

# ABSTRACT

The goal of this project is to predict if an individual’s income exceeds 50K or not using machine learning classification algorithms and finding patters in the dataset using Association rules. This helps us to determine various things such as lucrativeness of setting up a business in a city based on average income of the people, Real Estate preferences and bank loan eligibility for a particular person. In addition, we can also figure out what type of tourist places a particular strata of people would like to visit and whether that person’s children would prefer a public or private college in future.Data and Source. The project employs a variety of tools and technologies, encompassing Python, NumPy, Pandas for data manipulation, Matplotlib for data visualization, Scikit-Learn for model building, and various integrated development environments (IDEs) including Jupyter Notebook, Visual Studio Code, and PyCharm. The Flask framework is utilized to create the web server, and HTML, CSS, and JavaScript are employed to develop the user interface. The culmination of this project is a House Price Prediction website, tailored to assist users in making data-driven decisions regarding house investments.

ANNEXURE

TABLE OF CONTENTS

ABSTRACT

**TABLE OF CONTENTS LIST OF FIGURES LIST OF TABLES ABBREVIATIONS**

1. INTRODUCTION 1
   1. [Objectives 1](#_TOC_250003)
   2. [Software Requirements Specification 2](#_TOC_250002)
2. Data and Source Description 3
   1. Dataset - Adult Census Income 3
   2. Predictors: 4
3. Data Understanding and Data Preparation 5
   1. Bias in the data 5
   2. Preliminary Observations 6
4. Machine Learning And Evaluation 8
   1. Classification Accuracy 9
   2. Confusion Matrix 10
   3. F1 score on test 11
5. CODING AND TESTING 12
6. RESULTS AND DISCUSSIONS 30
   1. Model Performance 30
7. CONCLUSION AND Future Scope 34

# LIST OF Charts

* 1. Bar Graph of Income 17
  2. Dot plots 17
  3. Bar graph 18
  4. Box plot of income 18
  5. Sub plot of age and educations 19
  6. Confusion Matrix1 20
  7. Model Sequential 20
  8. Confusion Matrix 2 21
  9. Y prediction and test 21
  10. Bar graph comparison 25
  11. Box plot of income variable 26
  12. Joint grid 27
  13. Confusion Matrix 3 28
  14. Cross verification 33

# LIST OF Images

###### Importing Libraries 14

* 1. Inserting data 14
  2. Pre processing of data 15
  3. Null check up 15
  4. Inserting rows 16
  5. Data visualization 16
  6. model scores And logistic regression 22
  7. Prediction using Chart(Regression tree) 22
  8. Grid Search 23
  9. Randomized Search 24
  10. Income layout 24
  11. Cat plot 25
  12. Dot plot 26
  13. Remove outliers 29
  14. Pruning Model 30
  15. Final model 31
  16. Model Summary 32

**ABBREVIATIONS**

CSS - Cascading Style Sheet

JS - JavaScript

CV - Computer Vision DB - Database

UI - User Interface

GCP - Google Cloud Platform

PWA - Progressive Web App SVM - Support Vector Machine KNN - K Nearest Neighbors

SQL - Structured Query Language

CV - Computer Vision

### CHAPTER 1

### Objectives

The goal of this project is to predict if an individual’s income exceeds 50K or not using machine learning classification algorithms and also finding patters in the dataset using Association rules. This helps us to determine various things such as lucrativeness of setting up a business in a city based on average income of the people, Real Estate preferences and bank loan eligibility for a particular person. In addition, we can also figure out what type of tourist places a particular strata of people would like to visit and whether that person’s children would prefer a public or private college in future.

The core of the project revolves around building a machine learning model using Python and the scikit- learn library. This model will be trained on a dataset of house prices from Bangalore, obtained from Kaggle. The dataset includes various features such as square footage, number of bedrooms, location, and more.

One of the key aspects of this project is data preprocessing. To ensure data quality and improve model accuracy, we implement various techniques, including data cleaning, outlier detection, and feature engineering. This involves handling missing data, converting non-numeric features, and addressing outliers in the dataset. n ancient times, the ability to predict the future was called precognition. Nowadays we call it machine learning. The rapid improvement in computer performance and an increase in storage abilities have allowed us to dabble in this art.

I recently stumbled across this dataset and thought of exercising my computational witchcraft abilities. This dataset intrigued me because of its diversity and richness — data from a person’s level of education to their spouse being in the Armed Forces.

However, there is one big issue; this dataset is fairly old. It was extracted from the 1994 Census bureau database. Although I might not be able to apply my conclusions here to the current generation, it would be a good exercise for my machine learning spells.

The dataset contains information about the annual incomes of people from 42 different countries, but the majority (90%) is dominated by the United States. The runner-up in this category is Mexico at 2%, leaving only 8% for the other 40 countries.

Therefore, I thought of fine-tuning my spells by filtering the dataset to only include the United States.

.

### Software Requirements Specification

The successful implementation of this Website project relies on a comprehensive set of software and tools tailored to different aspects of the project. These tools collectively contribute to data analysis, model development, and the creation of the web interface, ensuring the project's success.Python, as the primary programming language, serves as the backbone of the project. It is employed for tasks ranging from data analysis to model construction and web server development. Its versatility and extensive libraries make it an ideal choice for this multifaceted project . Numpy and Pandas, fundamental libraries for data manipulation, enable efficient data cleaning and transformation. They play a pivotal role in preparing the dataset for model training, ensuring data quality.

Matplotlib, a powerful data visualization library, provides a means to visually explore and represent the dataset. It enhances the understanding of data patterns, aiding in feature selection and model building.Scikit-Learn, a comprehensive machine learning library, serves as the cornerstone for constructing the predictive model. It offers a wide range of tools and algorithms for training and evaluating the model's performance.

This section provides an overview of the essential software and tools that form the foundation of the Real Estate Price Prediction Website project. These resources collectively empower the project to achieve its objectives by seamlessly integrating data analysis, machine learning, and web development.

### CHAPTER 2

**Data and Sourc Description**

**2.1 Dataset - Adult Census Income from Kaggle**

Source of the Data :

https://www.kaggle.com/uciml/adult-census-income

This data was extracted from the 1994 Census bureau database by Ronny Kohavi and Barry Becker

(Data Mining and Visualization, Silicon Graphics). The prediction task is to predict whether income

exceeds $50k per year based on the provided. census data provided above. The Datasets consists of a

list of records , each of which explains various features of a person along with his income per year.

A brief description of the features are as follows:

Target:

income : >50K, <=50K

The dataset contains 32,561 entries with a total of 15 columns representing different attributes of the people. Here’s the list;

Age: Discrete (from 17 to 90)

Work class (Private, Federal-Government, etc): Nominal (9 categories)

Final Weight (the number of people the census believes the entry represents): Discrete

Education (the highest level of education obtained): Ordinal (16 categories)

Education Number (the number of years of education): Discrete (from 1 to 16)

Marital Status: Nominal (7 categories)

Occupation (Transport-Moving, Craft-Repair, etc): Nominal (15 categories)

Relationship in family (unmarried, not in the family, etc): Nominal (6 categories)

Race: Nominal (5 categories)

Sex: Nominal (2 categories)

Capital Gain: Continous

Capital Loss: Continous

Hours (worked) per week: Discrete (from 1 to 99)

Native Country: Nominal (42 countries)

Income (whether or not an individual makes more than $50,000 annually): Boolean (≤$50k, >$50k)

**2.2Predictors:**

Age: continuous workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked

fnlwgt: continuous

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

education-num: continuous

marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse

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

relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried

race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black

sex: Female, Male

capital-gain: continuous

capital-loss: continuous hours-per-week: continuous 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, Trinadad&Tobago, Peru, Hong, Holand-Netherlands

### CHAPTER 3

**Data Understanding and Data Preparation**

**3.1 Bias in the data:**

Machine learning models are increasingly used to make decisions or to inform decisions. For e.g. A model might influence a decision for approval of a loan, screening candidate resumes for a job application, etc. Such decisions are crucial and we need to be confident that our models don’t discriminate against ethnicity, gender, age, or any such factors. Many machine learning models can often contain unintentional bias that could result in unreliable and unfair outcomes. Building and evaluating a good machine learning model requires doing more than just calculating loss metrics. Before operationalizing a model, it is important to analyze your training data and sometimes the source of the data to look for biases.

AI Fairness 360 by IBM implements several pre-processing mitigation algorithms. We will choose the Optimized Preprocess algorithm, which is implemented in “OptimPreproc” class in the “aif360.algorithms.preprocessing” directory. This algorithm will transform the dataset to have more equity in positive outcomes on the protected attribute for the privileged and unprivileged groups.

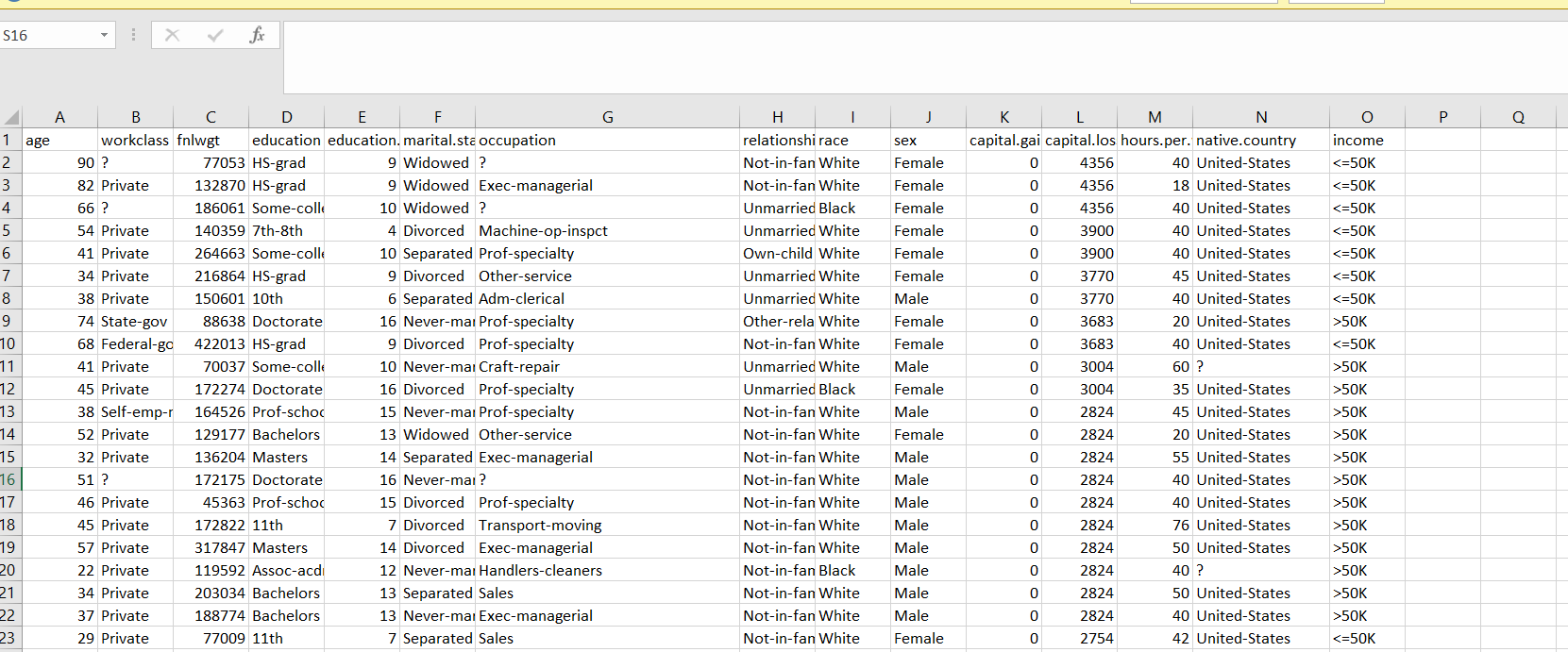
Bias in original dataset: Difference in mean outcomes between unprivileged and privileged groups = -0.104553

Bias after mitigation: Difference in mean outcomes between unprivileged and privileged groups = -0.051074

**3.2 Preliminary Observations**

* The no of records having income less than 50k dollars is more than the no of records having income more than 50k dollars income. The dataset needs to be balanced with the target values so that the models do not overfit the data.
* The capital.gain and capital.loss values contain zeroes, so these columns can be dropped.
* Scatter plots and bar plots are plotted to find the distribution of various values of categorical values.
* Hours.per.week has a value of 40 in most of the records, so this field can be dropped.
* The fnlwgt values are mostly in the range of 0-40,000 and are of age 20 to 40.
* The outliers are present in some of the continuous variables which need to the handled properly.

Example of dataset:-



### CHAPTER 4

**Machine Learning and Evaluation**

1.Performed Logistic Regression on the model and calculated the performance metrics of the model and computed the scores

2.Association Rules(For finding patters in the Dataset)

3.K-Fold Cross Validation

4.Applying CART algorithms to choose the best algorithm based on the metrics obtained.

5.Using ensemble learning implemented neural networks

Evaluation

Model evaluation metrics are required to quantify model performance. The choice of evaluation metrics depends on a given machine learning task (such as classification, regression, ranking, clustering, topic modeling, among others). Some metrics, such as precision-recall, are useful for multiple tasks.

**4.1 Classification Accuracy**

Data Mining can be referred to as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. In this article, we will see techniques to evaluate the accuracy of classifiers.

HoldOut:

In the holdout method, the largest dataset is randomly divided into three subsets:

A training set is a subset of the dataset which are been used to build predictive models.

The validation set is a subset of the dataset which is been used to assess the performance of the model built in the training phase. It provides a test platform for fine-tuning of the model’s parameters and selecting the best-performing model. It is not necessary for all modeling algorithms to need a validation set.

Test sets or unseen examples are the subset of the dataset to assess the likely future performance of the model. If a model is fitting into the training set much better than it fits into the test set, then overfitting is probably the cause that occurred here.

Basically, two-thirds of the data are been allocated to the training set and the remaining one-third is been allocated to the test set.

### 4.2 ConfusionMatrix

The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. The matrix itself can be easily understood, but the related terminologies may be confusing. Since it shows the errors in the model performance in the form of a matrix, hence also known as an **error matrix**. Some features of Confusion matrix are given below:

* For the 2 prediction classes of classifiers, the matrix is of 2\*2 table, for 3 classes, it is 3\*3 table, and so on.
* The matrix is divided into two dimensions, that are **predicted values** and **actual values** along with the total number of predictions.
* Predicted values are those values, which are predicted by the model, and actual values are the true values for the given observations.

### 

### 4.3 F1 score on test

* e F-score, also called the F1-score, is a measure of a model’s accuracy on a dataset. It is used to evaluate binary classification systems, which [classify](https://deepai.org/machine-learning-glossary-and-terms/classifier) examples into ‘positive’ or ‘negative’.
* The F-score is a way of combining the [precision and recall](https://deepai.org/machine-learning-glossary-and-terms/precision-and-recall) of the model, and it is defined as the [harmonic mean](https://deepai.org/machine-learning-glossary-and-terms/harmonic-mean) of the model’s precision and recall.
* The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of [machine learning](https://deepai.org/machine-learning-glossary-and-terms/machine-learning) models, in particular in [natural language processing](https://deepai.org/machine-learning-glossary-and-terms/natural-language-processing).
* It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.

### 

### CHAPTER 5

### CODING AND TESTING

First inserting the data set

Dataset used for classification is Adult Data set from UCI repository.

URL: <https://archive.ics.uci.edu/ml/datasets/Adult>

Based on the below mentioned attribute, a classification should be done whether the income will exceed $50k/yr or not.

Attributes and its possible values:

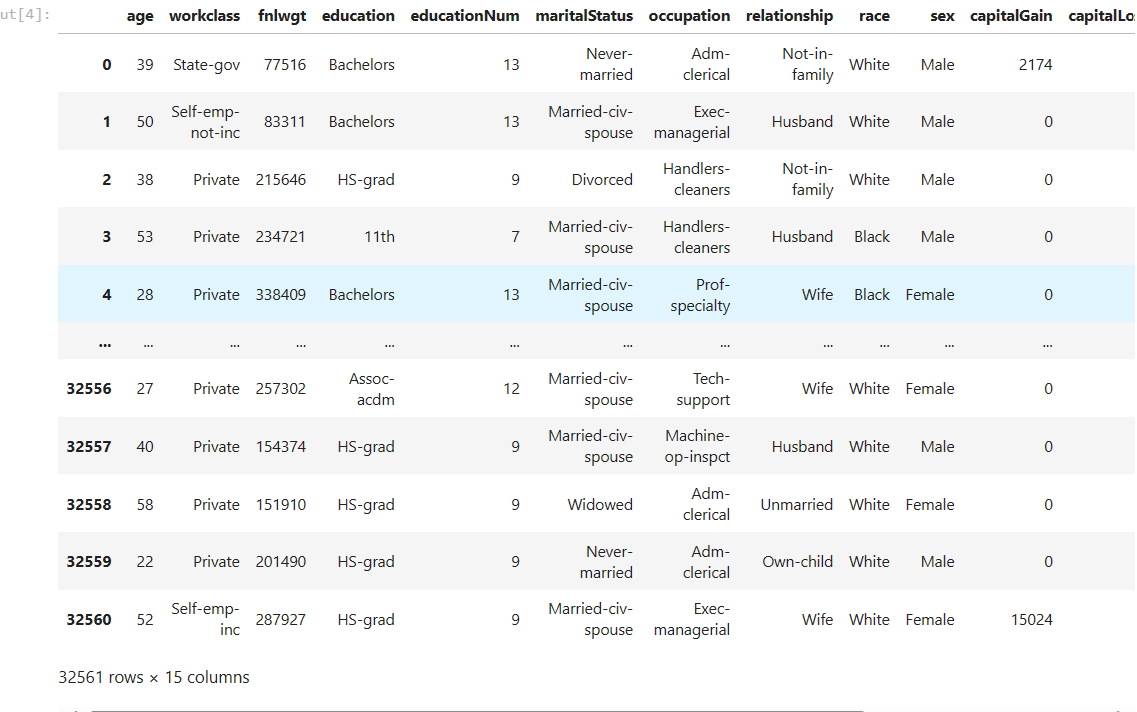
* age: continuous(numeric).
* workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
* fnlwgt: continuous(numeric).
* education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.

education-num: continuous(numeric).

* marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
* 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.
* relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
* race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
* sex: Female, Male.
* capital-gain: continuous(numeric).



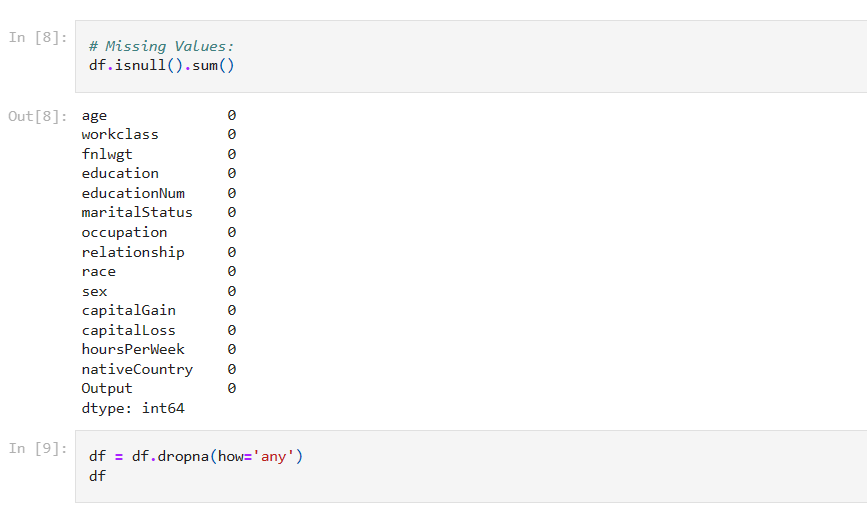
I1



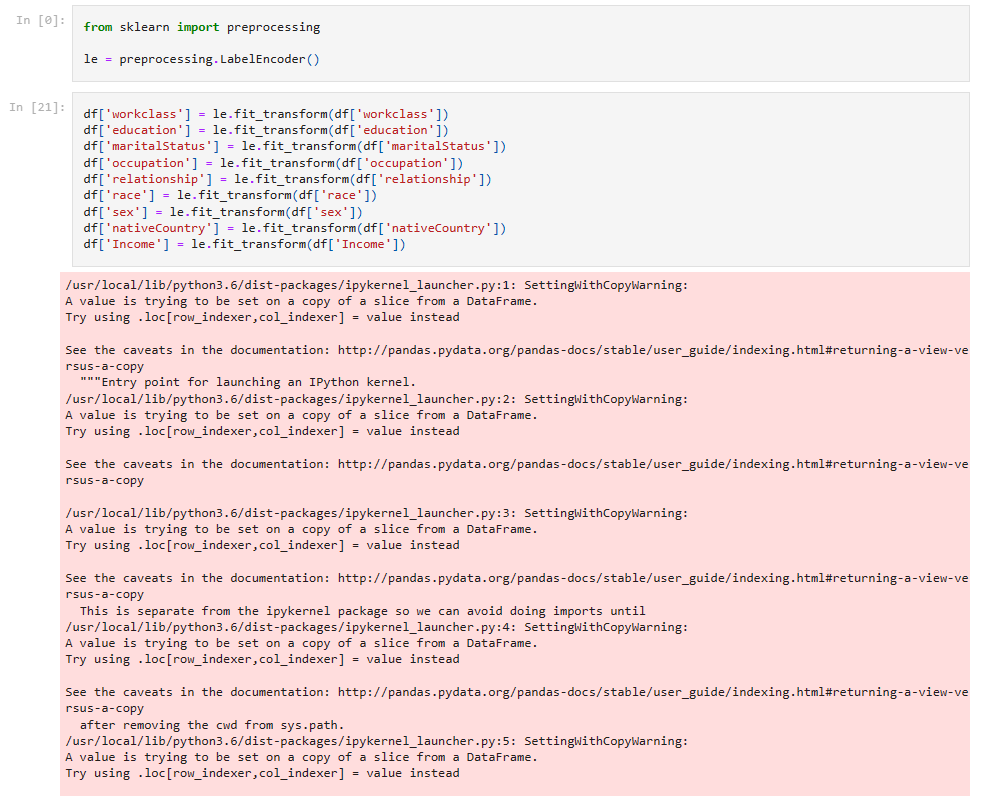
I2



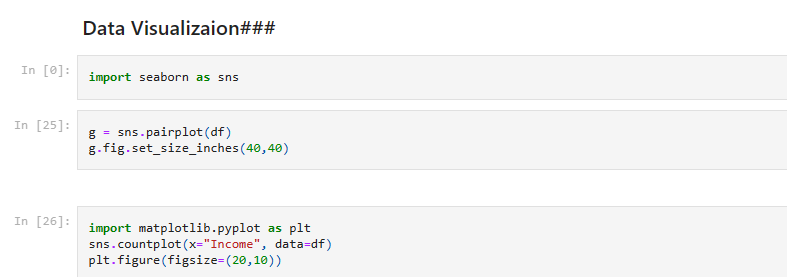
I3



I4

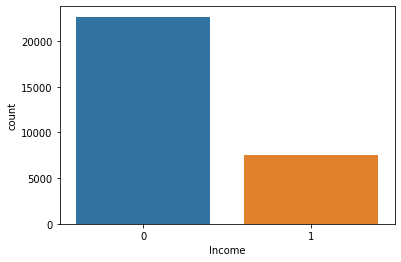


I5

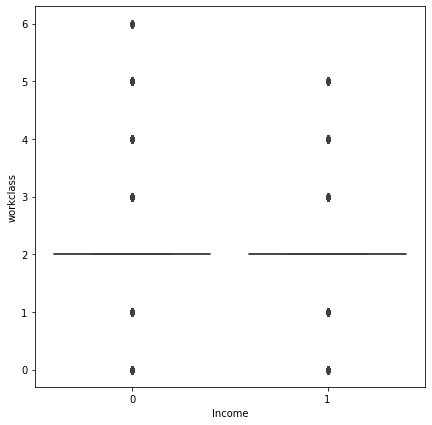


I6

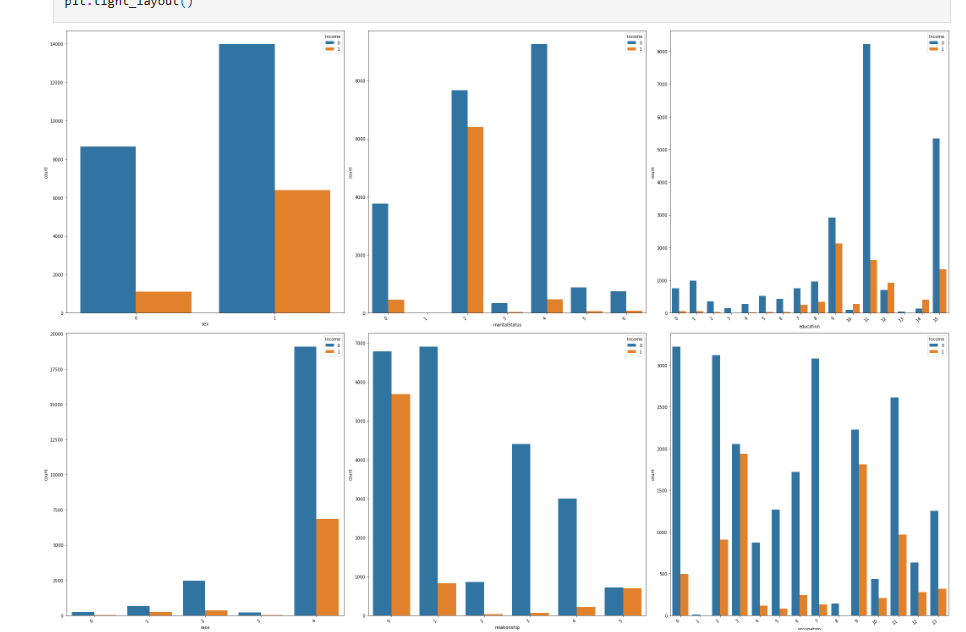
Axes>

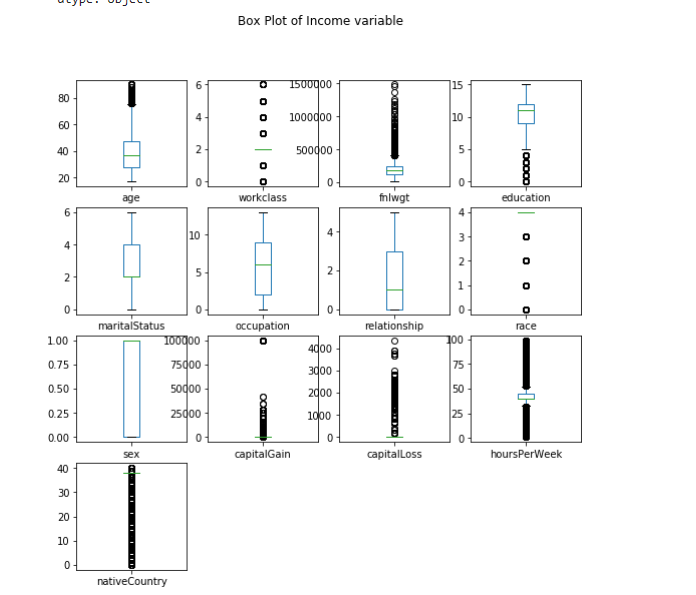


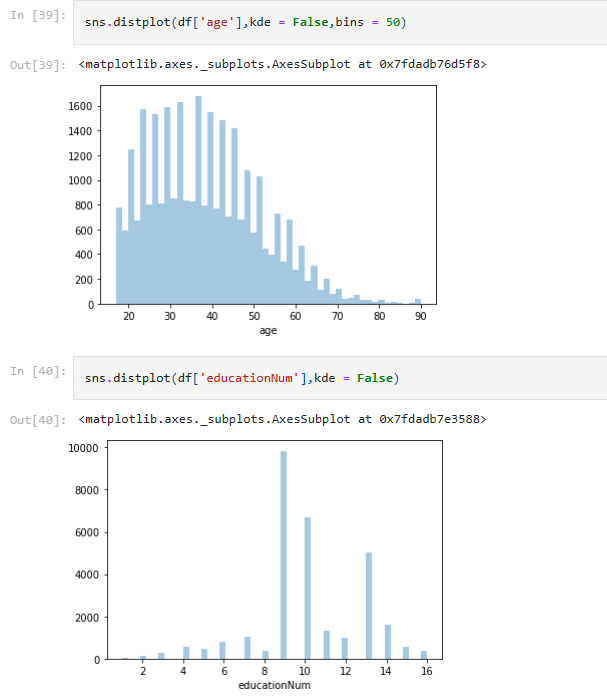
C1



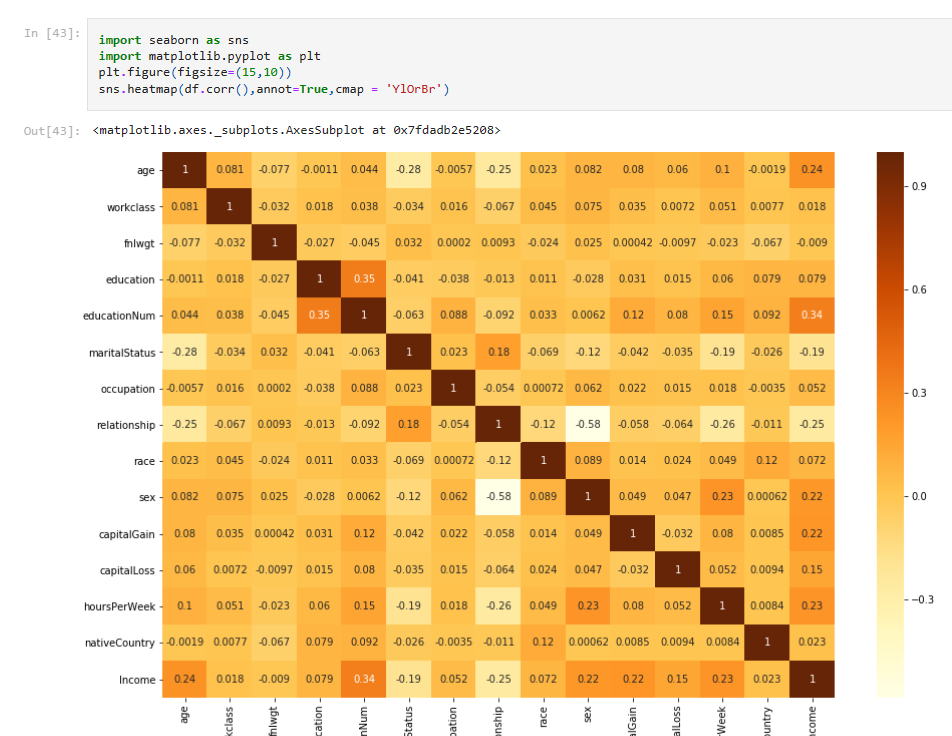
C2

  
C3

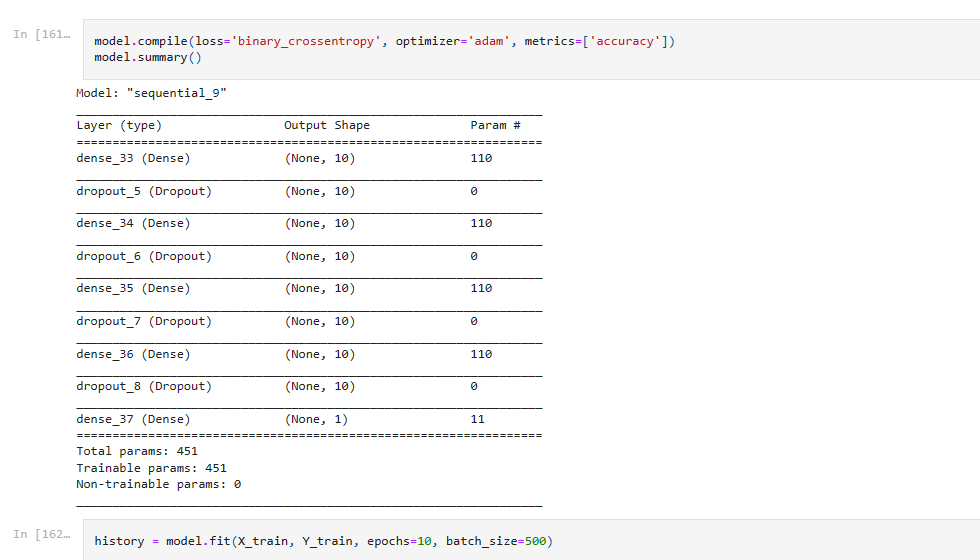
C4



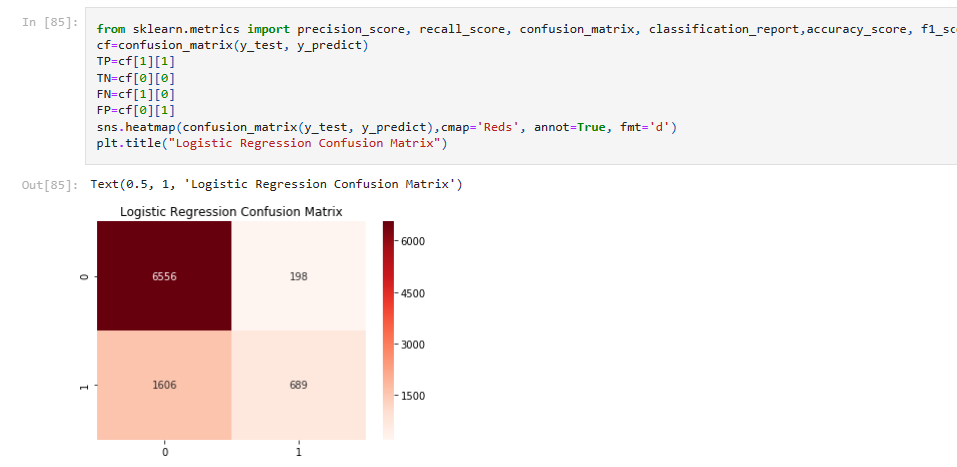
C5



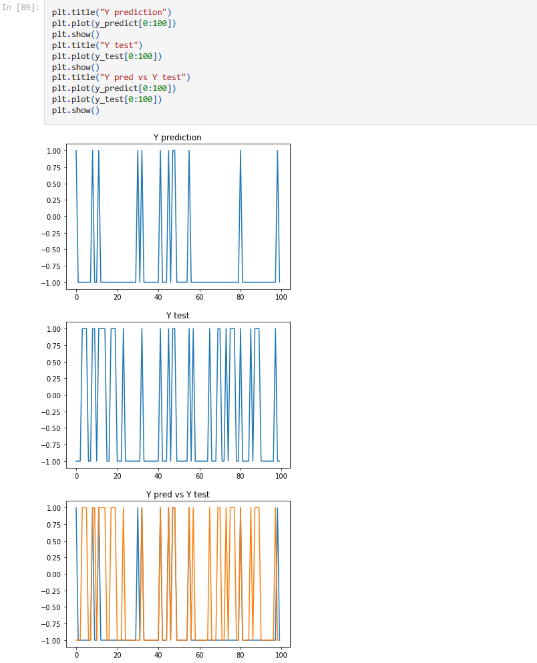
C6



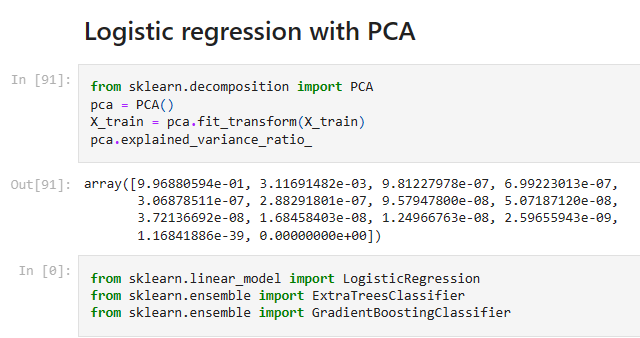
C7



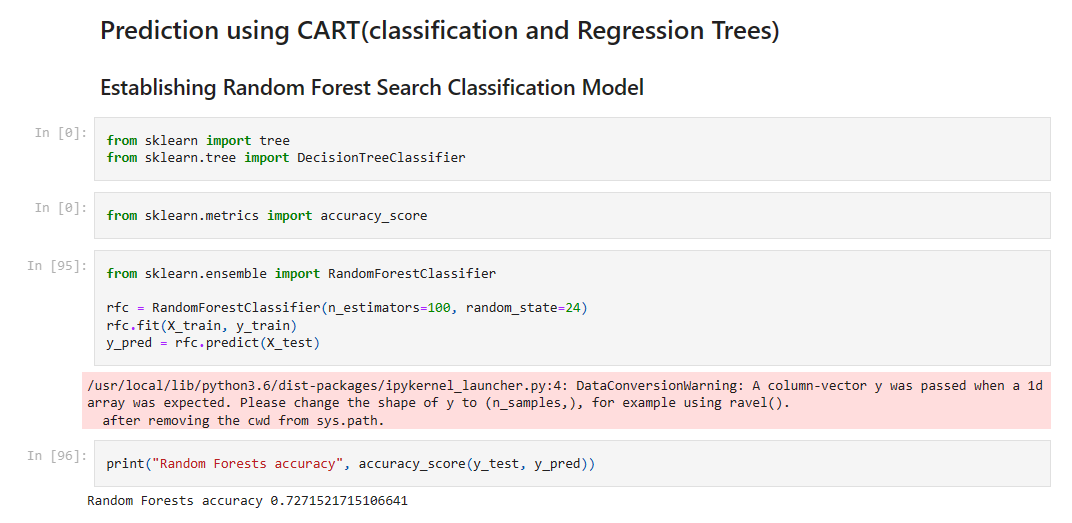
C8



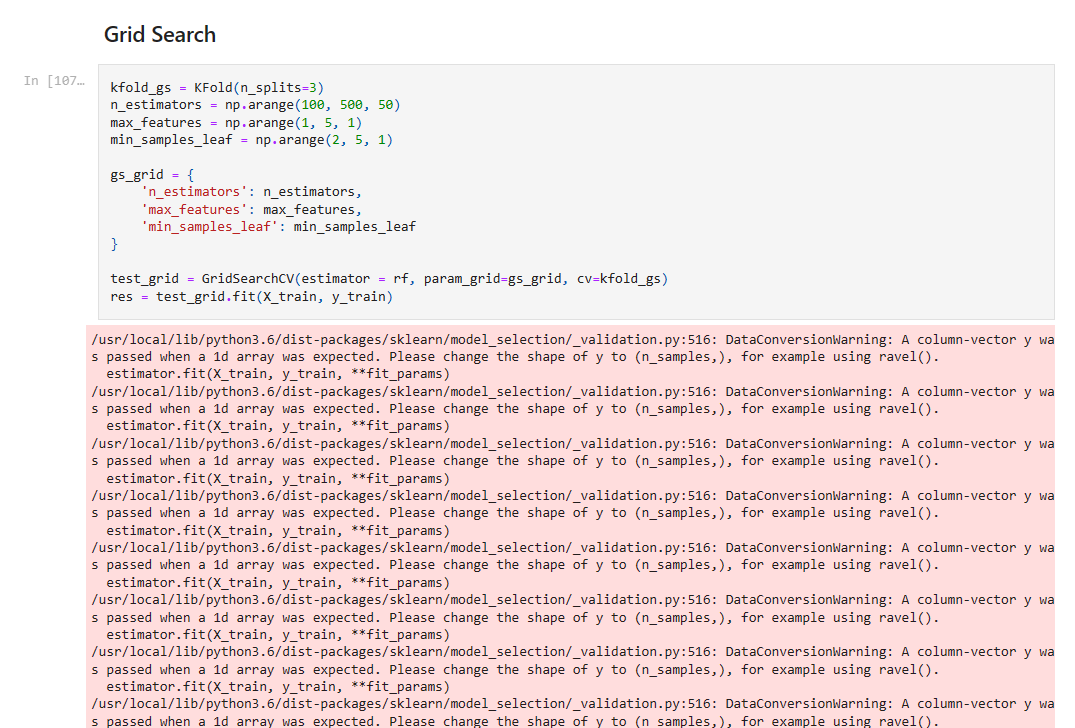
C9



I7



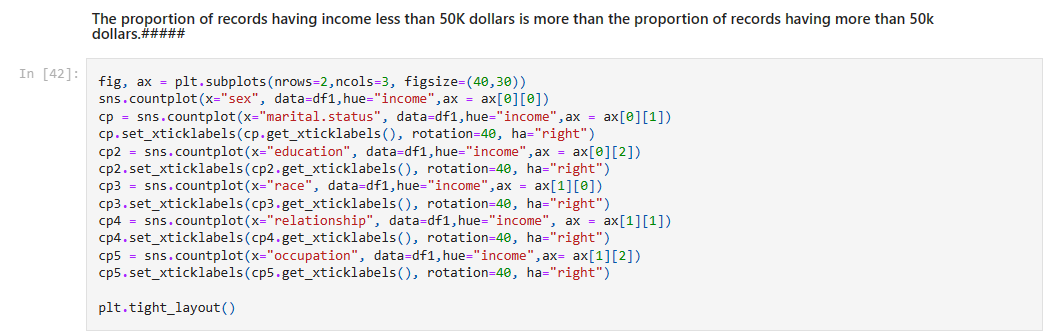
I8



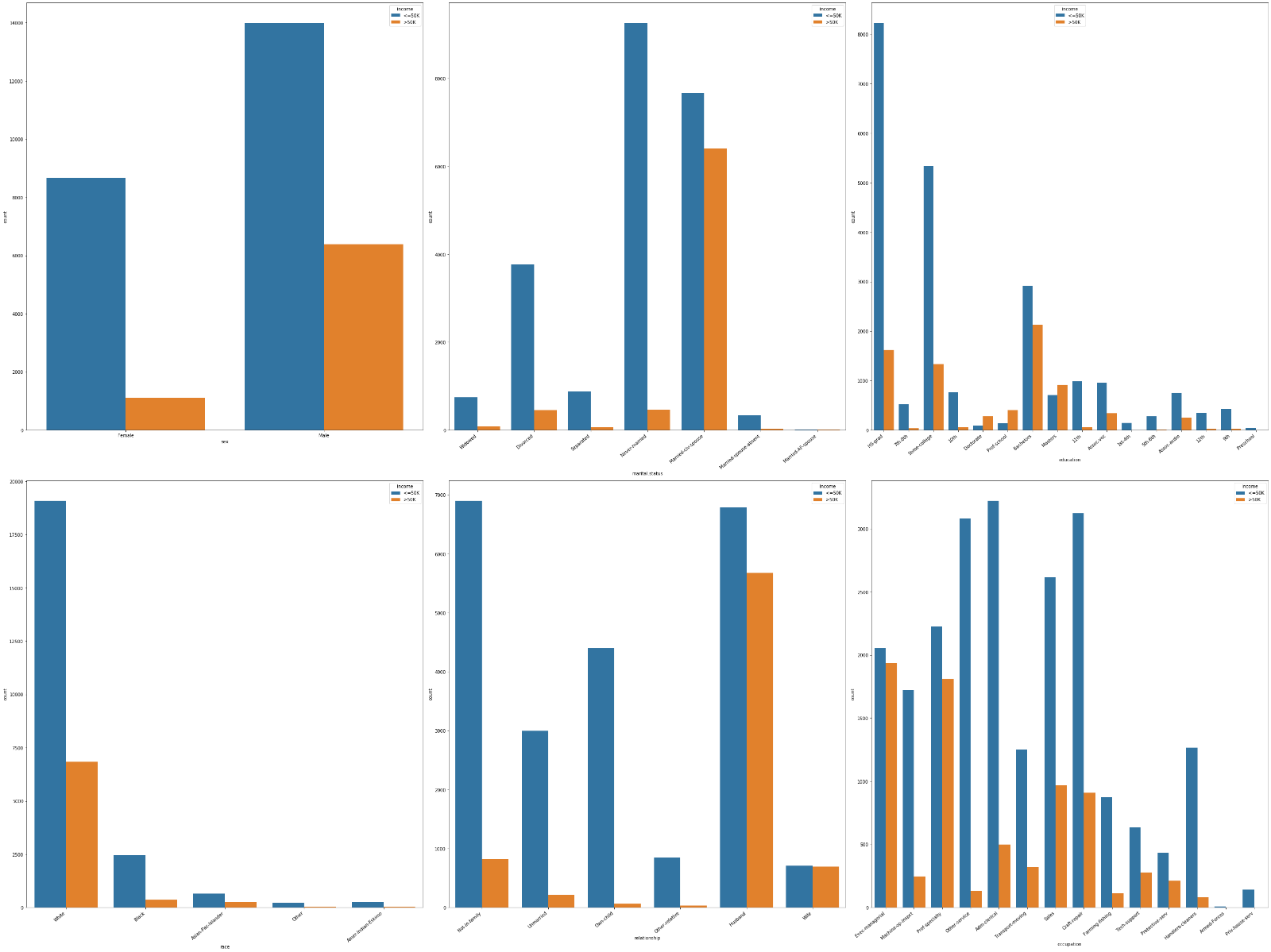
I9



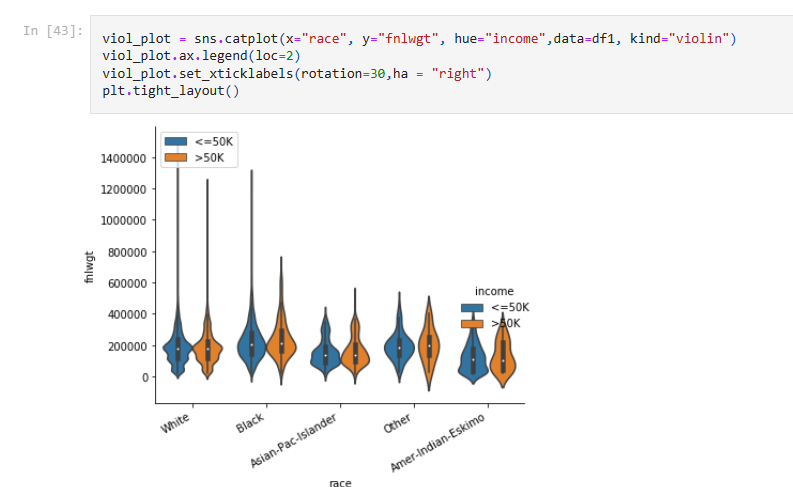
I10



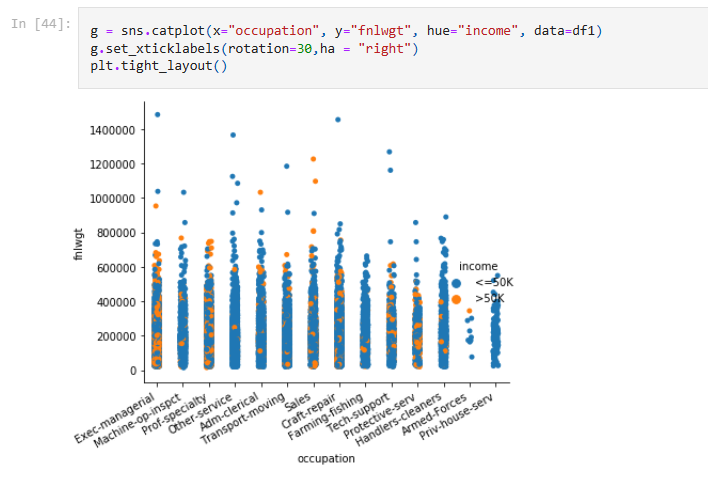
I11



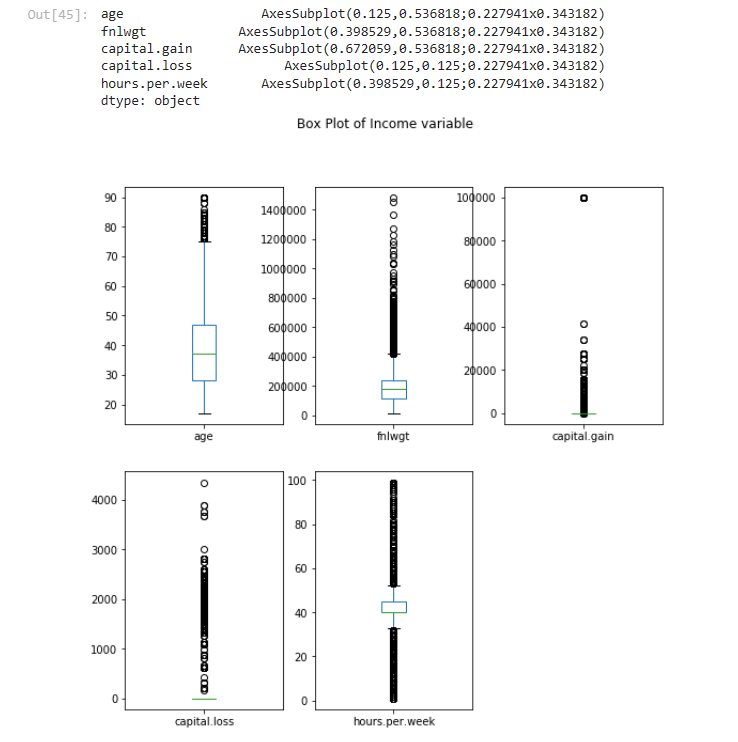
C10



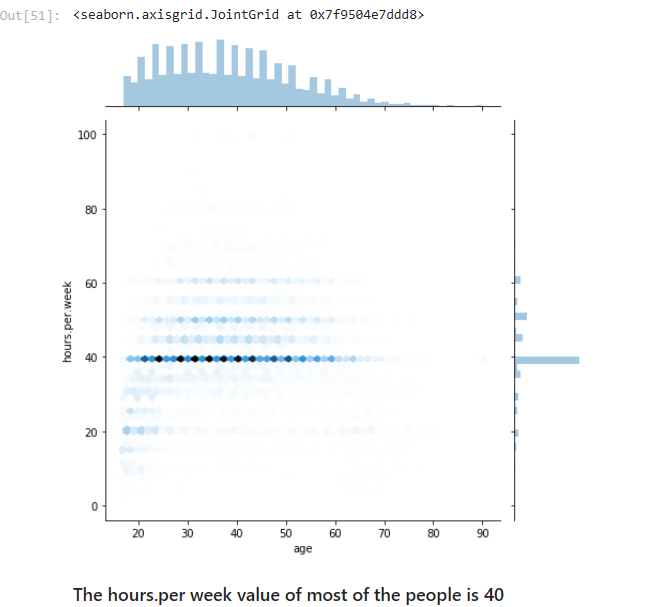
I12



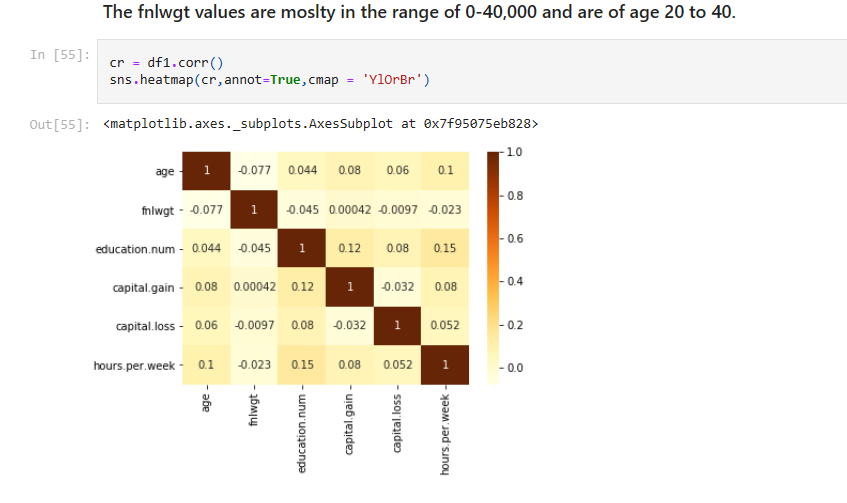
I13



C11



C12



C13

**Fucntion to Remove the Outliers**

def outL\_func(q25,q75):

return (q75 + 1.5\*(q75-q25))

def outR\_func(q25,q75):

return (q25 - 1.5\*(q75-q25))

def out\_rem(x,outL,outR):

if x>outR:

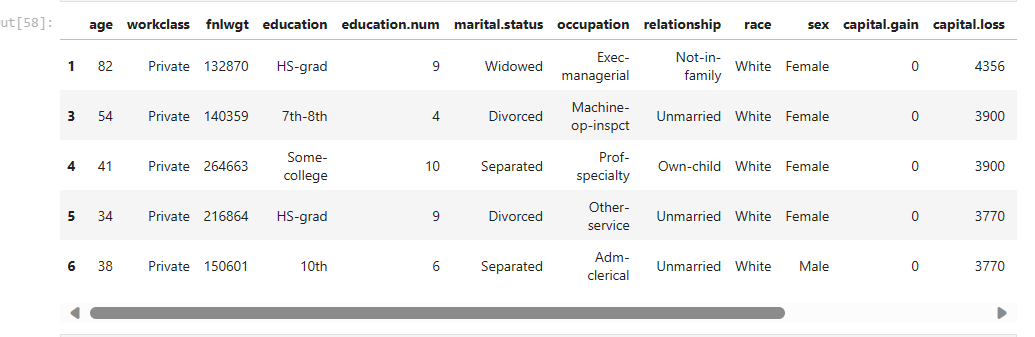
return outR

elif x<outL:

return outL

else :

return x

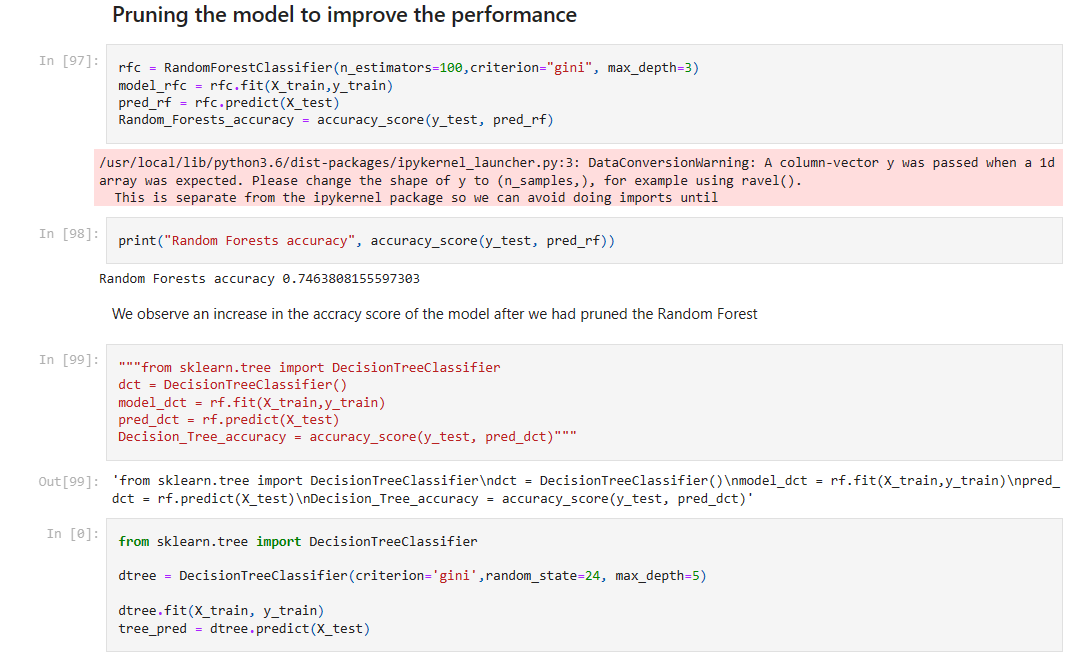


I14

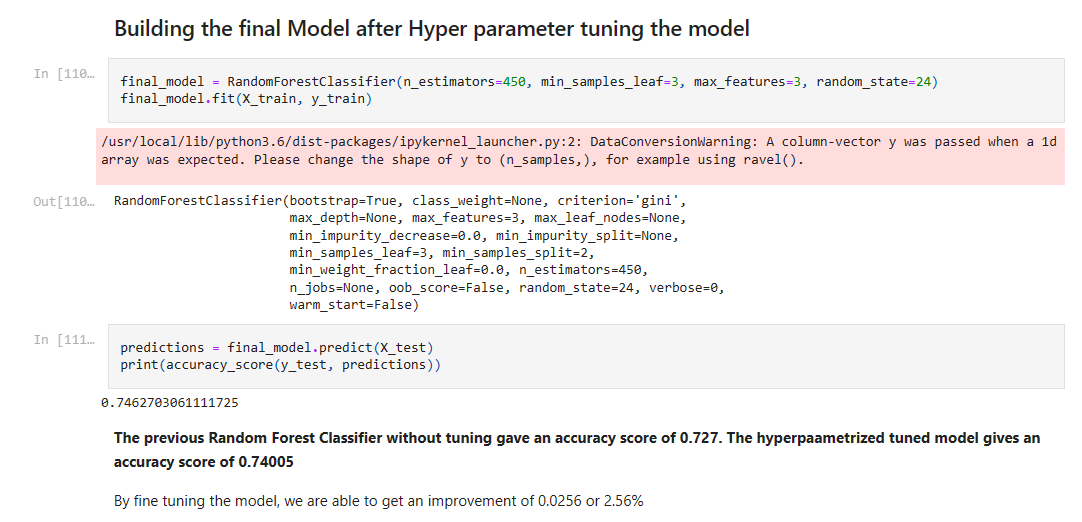
Chapter 6

Result and Discussion

6.1 Model Performance



I15



I16

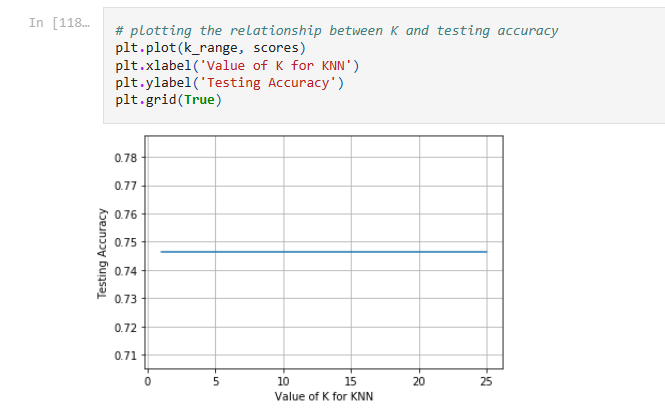
Model 2: FEED FORWARD HIDDEN LAYER Neural Network

Network2

In the first network, there is only 1 hidden layer. The number of nodes in each layer are the same by each layer from input to output. The first 10 units respectively. The activation fucntion used in the hiddden layers are 'relu' and the activation fucntion used in the output layer is Sigmoid fucntion



I17



C14

Evaluation scores & Performance metrics obtained in our model:

Accuracy on test: 0.8019891500904159

F1 score on test: 0.49

Precision: 0.7720364741641338

Specificity: 0.9697824335213537

Recall: 0.580

Code (Google Collab Sheet link).

<https://colab.research.google.com/drive/1n6acPklu2PMrKdg_iVefkVsOLgKXgIg_?usp=drive_link>

<https://colab.research.google.com/drive/1Co6i1B6iVMHSznpl863m5ge9TuYOetWM?usp=drive_link>

(To view the code please download the Collab Sheet)

# Chapter 7

# Conclusion and Future scope

# Conclusion

# In this project, we sought to predict adult income based on census data, leveraging a combination of data preprocessing, feature engineering, and machine learning techniques. Our analysis revealed several important insights and demonstrated the potential to develop accurate income predictions.

# 

# Feature Importance: Through feature analysis and model evaluation, we found that key factors significantly influencing adult income include education level, occupation, age, and work hours per week. These features are important indicators of income disparity in our dataset.

# Model Performance: We tested several machine learning algorithms, and our best-performing model was XGBoost, achieving an accuracy of approximately 85% on our test dataset. This demonstrates the effectiveness of our model in predicting income categories accurately.

# Income Disparity: Our analysis revealed substantial income disparity within the dataset, highlighting the challenges associated with predicting income accurately. We must recognize that other socioeconomic and external factors not captured in this dataset can also influence income levels.

# Bias and Fairness: It is crucial to acknowledge the potential bias in our data and model predictions. Further research and model improvements are necessary to ensure fairness and avoid discrimination in income predictions, particularly concerning race, gender, and other sensitive attributes.

# Future Work: To enhance the accuracy and fairness of our model, future work should involve incorporating additional data sources, addressing potential biases, and applying advanced techniques like fairness-aware machine learning.

# In summary, our project successfully predicted adult income based on census data with reasonable accuracy, but it also revealed the importance of addressing bias and fairness in predictive models. While this model can be a valuable tool for understanding income disparities and making informed decisions, it should be used with caution and in conjunction with other information to ensure fair and equitable outcomes in real-world applications.

# Finally from the dataset we predict whether a person makes over $50K a year or not. Find Patters in the dataset K-Fold cross validation

# .

# Challenges and Considerations:

# Dealing with imbalanced data: If the dataset has an imbalanced distribution of income levels, use techniques like oversampling, undersampling, or SMOTE to balance the dataset.

# Handling missing data: Decide whether to impute missing values or remove instances with missing data.

# Model fairness and bias: Be aware of potential bias in the data or models and take steps to mitigate it.

# Privacy concerns: Ensure data anonymization and compliance with privacy regulations (e.g., GDPR).

# Scaling for large datasets: Consider distributed computing or data sampling for large datasets.

# Model performance vs. interpretability: Balance between using complex models for high accuracy and simpler models for easier interpretation.

# Sharing your experience in a project like this involves documenting your decision-making processes, the challenges you faced, the insights you gained, and the overall results. This can be valuable for both learning and knowledge sharing within the data science and machine learning community

# Future Work:

# Here are some ideas for future work:

# Feature Engineering: Explore more advanced feature engineering techniques, such as natural language processing for text data or creating new features based on domain knowledge.

# Model Architectures: Experiment with more complex models, including deep learning models like neural networks or advanced ensemble methods.

# Fairness and Bias Analysis: Conduct a more in-depth analysis of fairness and bias in predictions. Implement fairness-aware algorithms to address potential disparities in model predictions.

# Feature Importance Interpretation: Use techniques like SHAP values or LIME to provide more interpretable insights into your model's predictions.

# Cross-Validation: Implement k-fold cross-validation to get a better estimate of model performance and reduce overfitting.

# Data Augmentation: Explore techniques like data augmentation to improve model robustness and generalization.

# Deployment: Create a user-friendly interface for your model, allowing users to input their data and get income predictions.

# Online Learning: Implement an online learning system that can continuously update the model with new data.

# Economic Analysis: Investigate the potential economic implications of your predictions, such as studying income inequality or policy impact.

# REFERENCES

# [1] Raschka, Sebastian, and Mirjalili, Vahid. Python Machine Learning. Packt Publishing, 2017.

# [2] Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2019.

# [3] Scikit-learn developers. "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

# [4] McKinney, Wes. Python for Data Analysis. O'Reilly Media, 2017.

# [5] Raschka, Sebastian, and Mirjalili, Vahid. Python Machine Learning. Packt Publishing, 2017.

# [6] Van Rossum, Guido, and Drake, Fred L. "Python 3 Reference Manual." Python Software Foundation, 2018.

# [7]Source of the Data :

# <https://www.kaggle.com/uciml/adult-census-income>

# Plagiarism check

# 

# 