PROJECT 3

Predicting Divorce: Decision Trees, Perceptrons, and Logistic Regression

Artificial Intelligence & Machine Learning

Table Of Contents

1. [Aim](#1-aim)
2. [Introduction](#2-introduction)
3. [Technical Requirements](#3-technical-requirements)

#### [Hardware Specifications](#hardware-specifications)

#### [Software Requirements and Environment Setup](#software-requirements-and-environment-s)

#### [Suggested Tools / Tech Stacks](#suggested-tools--tech-stacks)

1. [Modeling Approach](#4-modeling-approach)
2. [Implementation Steps](#5-implementation-steps)
3. [Expected Output](#6-expected-output)
4. [Conclusion](#7-conclusion)
5. [Reference](#8-reference)

**Submitted By:**  
**PRAKASH N**  
**E-mail ID:** prakashpoint2005@gmail.com

**1. Aim**

This project aims to develop predictive models using decision trees, perceptrons, and logistic regression to predict divorce based on various attributes.

**2. Introduction**

Predicting divorce can provide valuable insights for counseling and intervention. This project uses a dataset from the UCI Machine Learning Repository containing attributes related to personal relationships to build and evaluate models for predicting divorce.

**3. Technical Requirements**

**Hardware Specifications**

* **Processor:** Intel Core i5 or higher
* **Memory:** 8 GB RAM or higher
* **Storage:** 256 GB SSD or higher
* **Graphics:** Integrated graphics card sufficient for data visualization tasks

**Software Requirements and Environment Setup**

* **Operating System:** Windows 10, macOS, or Linux
* **Programming Language:** Python 3. x
* **Development Environment:**
  + **Primary IDE:** Visual Studio Code (VS Code) with Jupyter Notebook integration
  + **Alternative:** Jupyter Notebook standalone, Google Colab

**Libraries**

* **Data Handling and Analysis:** pandas, numpy
* **Data Visualization:** matplotlib, seaborn
* **Machine Learning:** sci-kit-learn

**VS Code Extensions**

* **Python Extension for Visual Studio Code:** Microsoft’s official extension for Python development
* **Jupyter Extension for Visual Studio Code:** Provides support for Jupyter Notebooks inside VS Code (Microsoft’s official extension for Jupyter Notebooks like Jupyter, Jupyter Keymap, Jupyter Power Toys)

**Suggested Tools / Tech Stacks**

1. Python Programming Language
2. Jupyter Notebook for interactive data analysis
3. Visual Studio Code (VS Code) as a versatile IDE with Jupyter support
4. Essential libraries for data handling, visualization, and machine learning
5. Git for version control and collaboration

**4. Modeling Approach**

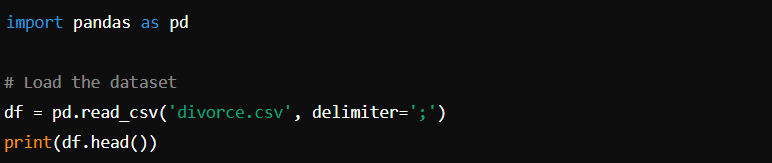
The project employs decision trees, perceptrons, and logistic regression to predict divorce based on various attributes. The approach involves the following steps:

1. Exploratory Data Analysis (EDA) to understand the data.
2. Splitting the data into training and testing sets.
3. Training the models on the training set.
4. Making predictions on the test set.
5. Evaluating the models' performance using appropriate metrics.

**5. Implementation Steps**

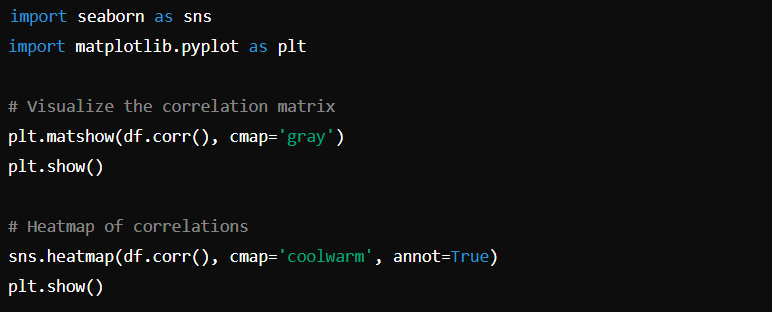
**Step 1: Load the Dataset**

First, we load the dataset using pandas, a powerful data manipulation and analysis library. The dataset contains various attributes that will be used for our analysis and model building.



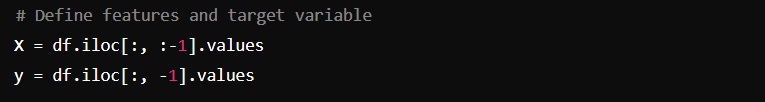
**Step 2: Data Visualization**

Next, we use seaborn and matplotlib to visualize the relationships between the attributes. Correlation matrices and heatmaps provide insights into the relationships between the variables.



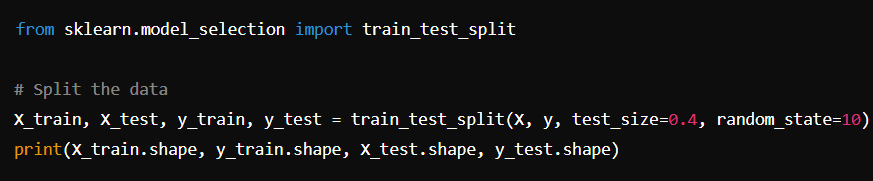
**Step 3: Define Features and Target Variable**

We then define the features (independent variables) and the target (dependent variable). This prepares the data for model training.



**Step 4: Split the Data into Training and Testing Sets**

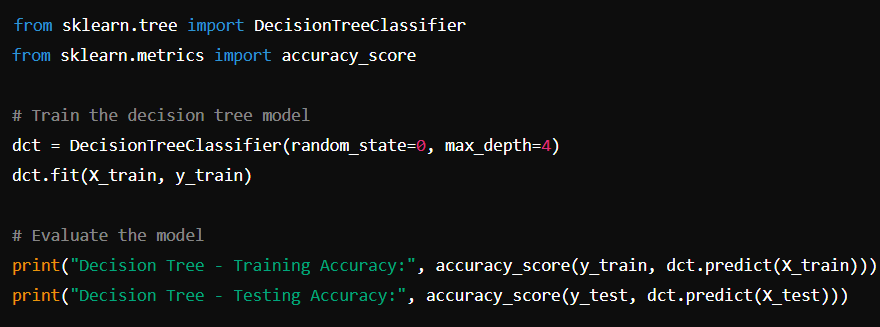
We split the dataset into training and testing sets to evaluate our models' performance. The training set will be used to train the models, while the test set will be used to assess how well the models generalize to unseen data.



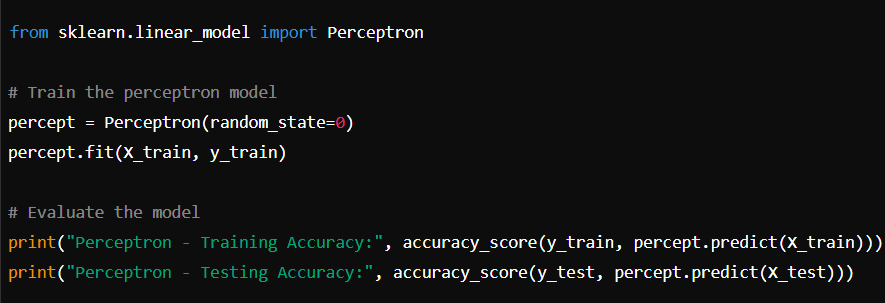
**Step 5: Train the Models**

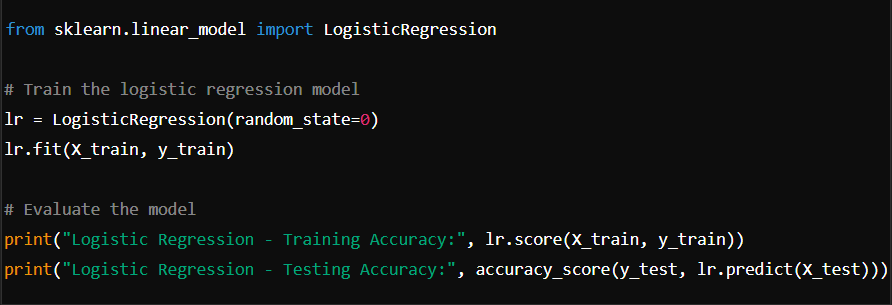
Using the training set, we train decision tree, perceptron, and logistic regression models. We then print the accuracy of each model on both the training and testing sets.

**Decision Tree**

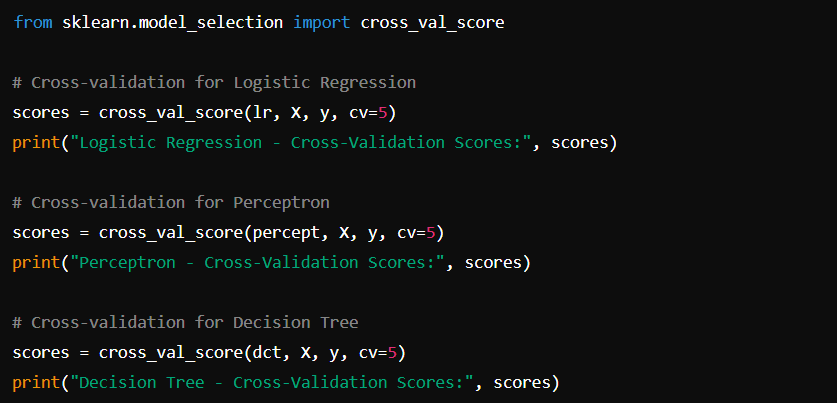


**Perceptron**

**Logistic Regression**

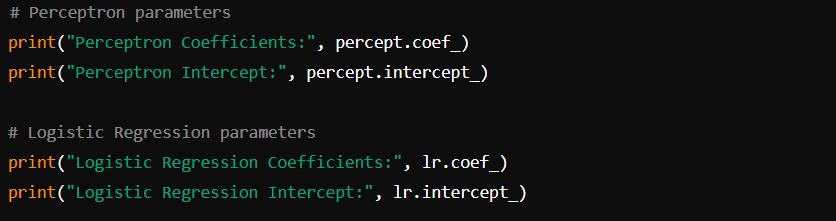
**Step 6: Cross-Validation**

To ensure the robustness of our models, we use cross-validation to evaluate them.



**Step 7: Evaluate Model Parameters**

We also examine the parameters of the trained models for further insights.



**6. Expected Output**

* Trained models capable of predicting divorce based on various attributes.
* Visualization of the data and correlation matrices.
* Evaluation metrics including training and testing accuracy for each model.
* Cross-validation scores to ensure model robustness.

**7. Conclusion**

This project demonstrates the use of decision trees, perceptrons, and logistic regression to predict divorce. By analyzing the relationships between various attributes and divorce, the models provide insights into factors that may contribute to marital dissolution. The evaluation metrics and cross-validation scores help in understanding the models' accuracy and reliability.

**8. Reference**

* UCI Machine Learning Repository: [Divorce Predictors Data Set](https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set)
* Scikit-learn documentation: <https://scikit-learn.org/stable/user_guide.html>
* Pandas documentation: <https://pandas.pydata.org/pandas-docs/stable/>
* Seaborn documentation: <https://seaborn.pydata.org/>
* **Divourse Observations Dataset**
* **Link:** https://drive.google.com/file/d/1qdfhN6ovNpo1qxOealuQatgWzpj-IcZk/view?usp=sharing
* **My Project Location**
* **Link:**https://drive.google.com/file/d/10tNWQW0Yr8mO9nxEFUqyjorVUHyTfz0d/view?usp=sharig