**Report**

**Maqsad:**

Bizga voyaga yetgan yoshlarning yillik daromadi haqida malumot berilgan. Maqsadimiz qaysiki shu voyaga yetgan yoshlarning yillik moashi 50.000$ dan kam yoki kopligini bashorat qilishdir. Biz buni yoshlarning demografik yokida ish joyiga qarab predict qilishimiz mumkin.

**Asosiy Natijalar:**

Bizga umumiy 13 ustunlik va ohirgi ustunini yani incomeni target variable qilb belgilab oldik hamda bizga 48841 qatorlik dataset berilgan shular yordamida ishladik.

Biz umumiy 4 model bilan bu datasetni korib chiqdik ular: Random Forest, Decision Tree, Linear Regression va KNN modellari. Har bir modelda har xil natijaga erishdik va natijaga erisha olmagan modellarimiz ham boldi

**Dataset Xusiyatlari Haqida Umumiy Malumot**

* **Numeric Ustunlar**:

age, capital-gain, capital-loss, hours-per-week

* **Categorik Ustunlar**:

workclass, education, marital-status, occupation, relationship, race, sex, native-country

**Target Variable**

* **Income**:

**Malumotlar Tahlili:**

Bizda null qiymatlar yoq ekan, Label encoding yordamida ishladik, standart scalingdan foydalandik va feature Engineeringda yangi ustun CapitalLoss\_to\_Hours , CapitalGain\_to\_Age, and CapitalLoss\_to\_Hours lar qoshib olindi

**Ishlatilgan Kutubxonalar Royxati**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

from sklearn.neighbors import KNeighborsRegressor

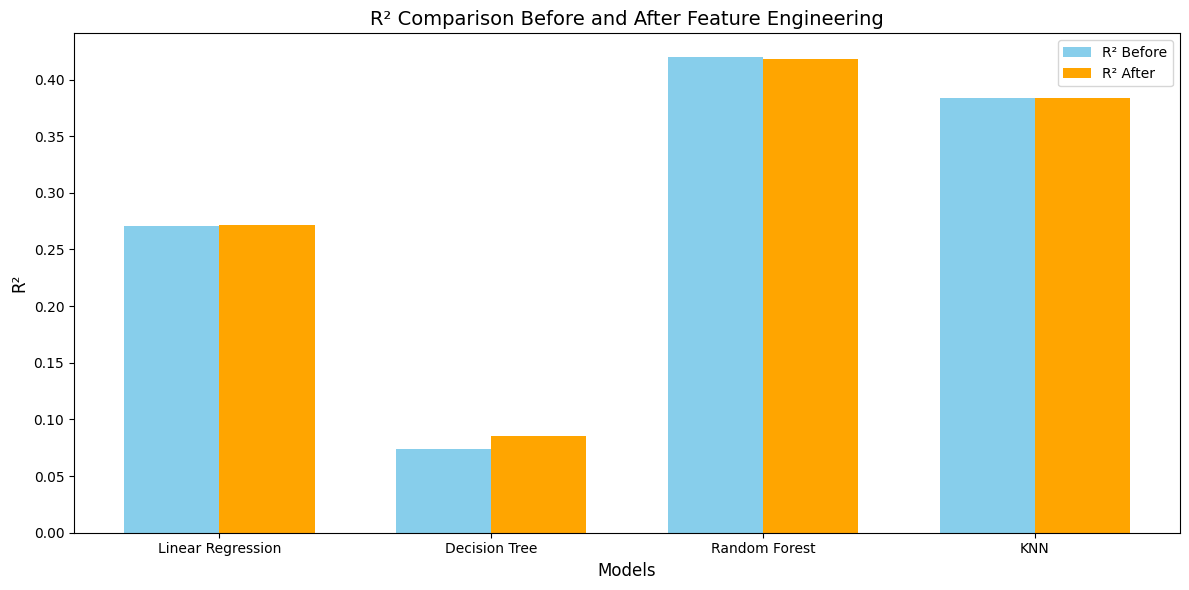
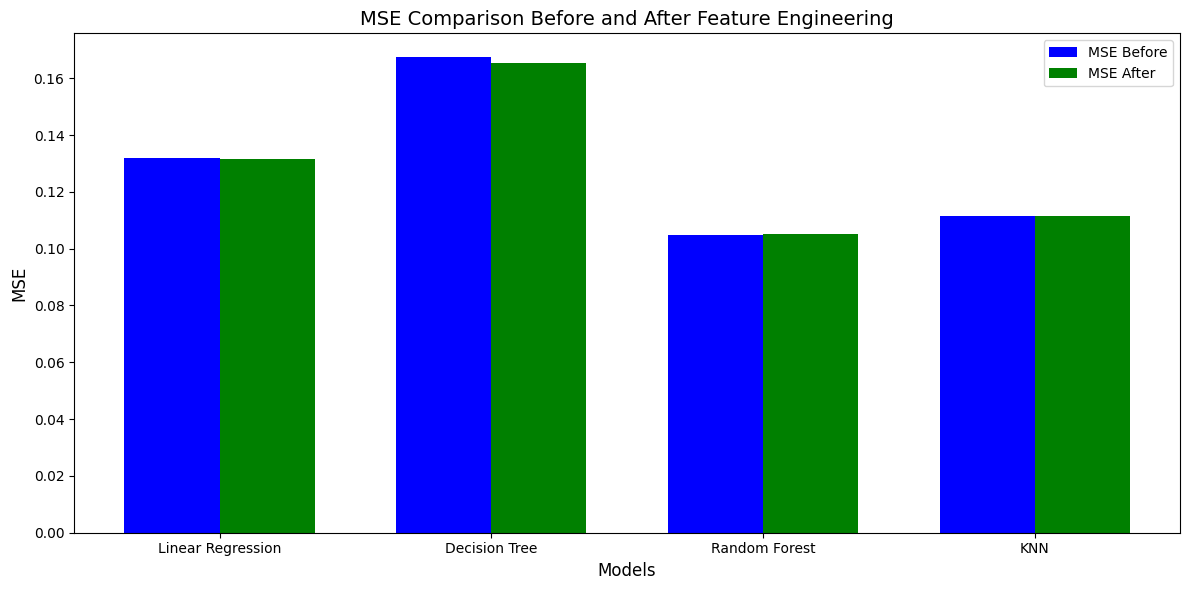
from sklearn.preprocessing import LabelEncoder,StandardScaler, OneHotEncoder

from sklearn.model\_selection import train\_test\_split, cross\_val\_score,cross\_val\_predict

from sklearn.metrics import r2\_score, mean\_squared\_error

from sklearn.model\_selection import RandomizedSearchCV,GridSearchCV

| **Model** | **Metric** | **Before** | **After** |
| --- | --- | --- | --- |
| **Linear Regression** | MSE | 0.13187964485376374 | 0.13164911799230644 |
|  | R² | 0.27055770892906383 | 0.27183278092490715 |
| **Decision Tree** | MSE | 0.16741843408684914 | 0.1653604066995601 |
|  | R² | 0.07398836065083247 | 0.0853715594313228 |
| **Random Forest** | MSE | 0.10489834247012175 | 0.1051794438236336 |
|  | R² | 0.41979456082251076 | 0.41823975518471723 |
| **KNN** | MSE | 0.11136890951276102 | 0.11136890951276102 |
|  | R² | 0.3840050706904695 | 0.3840050706904695 |



1. **Linear Regression**:

* **R²** ko'rsatkichi ozgina yaxshilandi (**0.2706** dan **0.2718** gacha) Feature Engineeringdan keyin.
* **MSE**da ham kichik yaxshilanish kuzatildi.

1. **Decision Tree**:

* Ham **MSE** **0.1654** gacha kamaydi, ham **R²** **0.0854** gacha oshdi sezilarli yaxshilandi.

1. **Random Forest**:

* **MSE** deyarli o'zgarmadi **~0.105**.
* **R²**da ozgina pasayish kuzatildi **0.4198** dan **0.4182** gacha, bu Feature Engineering katta ta'sir ko'rsatmaganini ko'rsatmoqda.

1. **KNN**:

* **MSE** va **R²** ko'rsatkichlari hech qanday o'zgarishga uchramadi