**Model Report**

**Introduction**

This report presents an analysis of machine learning models applied to the Adult Income dataset, aiming to predict whether an individual’s income exceeds $50K based on various features. We evaluate the models using metrics such as accuracy, R-squared (R²), mean squared error (MSE), and root mean squared error (RMSE).

**Key Results**

The dataset consists of 15 columns and 48,841 rows, with the last column as the target variable. We evaluated four models: Decision Tree, XGBoost, LightGBM, and KNN, each yielding different performance results.

**General Information About the Dataset Features**

* Numerical Features:

age, fnlwgt, educational-num, capital-gain, capital-loss, hours-per-week

* Categorical Features:

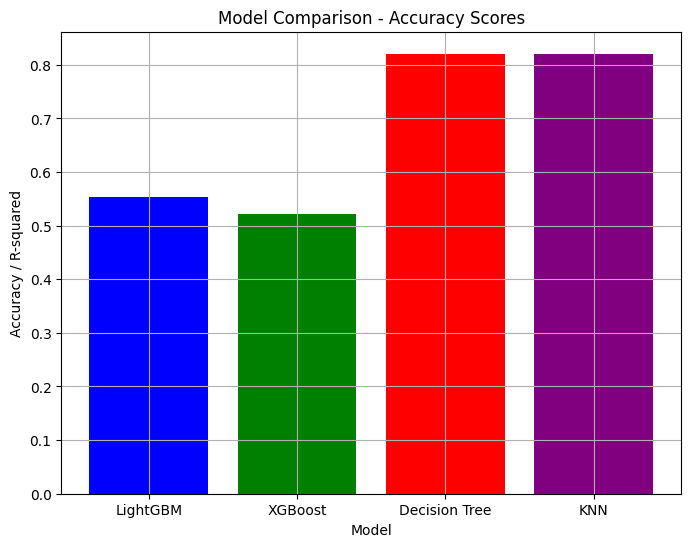
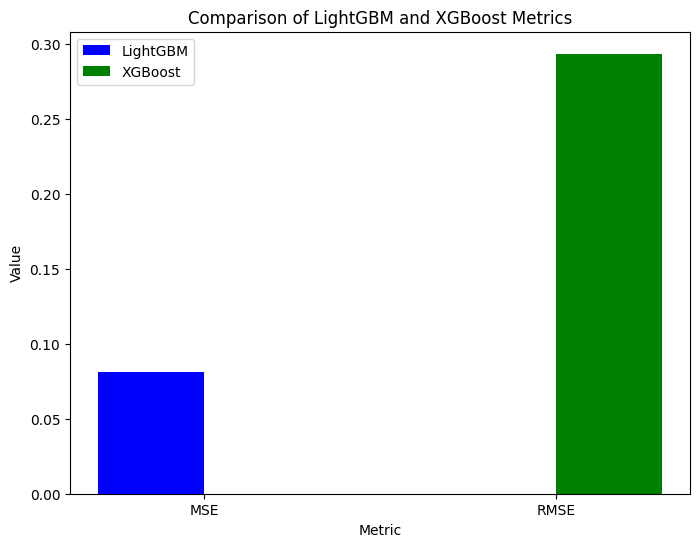
workclass, education, marital-status, occupation, relationship, race, gender, native-country, income

**Results and**  **Model Comparison**

We compare the following models:

* LightGBM (LGBM)
* XGBoost (XGB)
* Decision Tree
* K-Nearest Neighbors (KNN)

| Model | Metric | Value |
| --- | --- | --- |
| LightGBM | Mean Squared Error (MSE) | 0.0814 |
|  | R-squared (R²) | 0.5528 |
| XGBoost | Root Mean Squared Error (RMSE) | 0.2932 |
|  | Mean Absolute Error (MAE) | 0.1844 |
|  | R-squared (R²) | 0.5210 |
| Decision Tree | Test Accuracy | 0.8200 |
| KNN | Test Accuracy | 0.8207 |

****

**Model Analysis**

* **LightGBM** demonstrated a relatively low Mean Squared Error (MSE), indicating a good fit for the data. With an R-squared (R²) value of 0.5527, it shows moderate predictive power, capturing the relationships between the features and the target variable effectively.
* **XGBoost**, while exhibiting a slightly higher Root Mean Squared Error (RMSE) than LightGBM, still delivered reasonable performance. The Mean Absolute Error (MAE) is 0.1844, and an R² value of 0.5210
* **The Decision Tree** model achieved the highest accuracy among the models tested with accuracy score is 0.8199
* **K-Nearest Neighbors (KNN)** showed a very similar performance to the Decision Tree, with an accuracy of 0.8207,

**Data Analysis**

* **Null Values**

We confirmed that the dataset is **free of null values**

* **Encoding**

To handle the **categorical features**, we employed **Label Encoding**, transforming the categorical columns such as workclass, education, marital-status, occupation, relationship, race, gender, native-country, and income into numeric values

* **Scaling**

We applied **Standard Scaling** across all numerical variables.

* **Target Variable**

The target variable for this classification task is the **“income”** column, which indicates whether an individual earns more than $50K per year.

**List of Libraries Used**

* *import numpy as np*
* *import matplotlib.pyplot as plt*
* *import seaborn as sns*
* *from sklearn.model\_selection import train\_test\_split, GridSearchCV, RandomizedSearchCV*
* *from sklearn.preprocessing import LabelEncoder, StandardScaler, OneHotEncoder*
* *from sklearn.tree import DecisionTreeClassifier*
* *from sklearn.neighbors import KNeighborsClassifier*
* *from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix*

**Model Evaluation**

* Decision Tree and KNN Models: The accuracy metric was used to evaluate the performance of these models, which is ideal for classification tasks.
* XGBoost and LightGBM Models: We used R-Squared (R²) for evaluating these regression-based models.

**Model Endpoint**

* The completed models and code have been uploaded to the following GitHub repository for further review and utilization:

[**https://github.com/ShodiyAbdulloh/AI\_ShodiyAbdulloh**](https://github.com/ShodiyAbdulloh/AI_ShodiyAbdulloh)

**Conclusion**

* Decision Tree and KNN Models: Both models demonstrated similar performance in terms of accuracy. Given their simplicity and ease of interpretation, they are strong candidates for this dataset.
* LightGBM and XGBoost: These sophisticated models showed lower MSE values, indicating better generalization and predictive power. While LightGBM outperformed XGBoost in terms of MSE, both models still offer competitive results.