

Data Science for Social Good

ABINAY SHREE J
M AKSHAYA



UN SDG Goal

- PEACE AND JUSTICE
- ABINAY SHRI J, M AKSHAYA
- URK20CS2117, URK20CS2096
- abinayshri@karunya.edu.in, makshaya@karunya.edu.in
- <https://github.com/Akshaya1aks/AI-HACKATHON>

TABLE OF CONTENTS

- Problem Statement And Mapping With UN SDG
- Solution Found And Dataset Used
- Data Analysis And Model Building
- Insights Gained And References Used

PROBLEM STATEMENT

- The objective is to predict the number of Child Care Institutions and Juvenile Justice Boards for the future years based on the historical data.
- Based on the given problem statement, the relevant UN Sustainable Development Goal (SDG) that matches this problem is SDG 16 - Peace and Justice.
- SDG 16 aims to promote peaceful and inclusive societies, provide access to justice for all, and build effective, accountable and inclusive institutions at all levels. This aligns with the problem statement of creating a model to predict the future of State/UT-wise Child Care Institutions and Juvenile Justice Boards, which can help ensure the proper functioning of institutions and access to justice for children in need.

SOLUTION

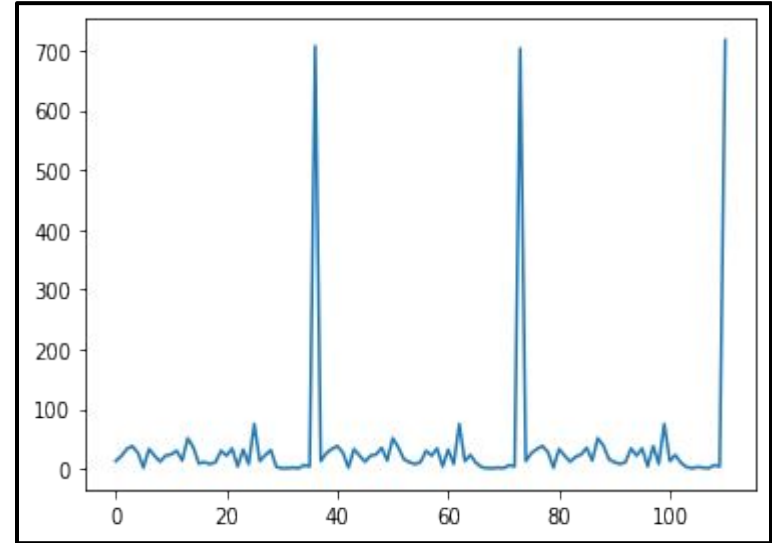
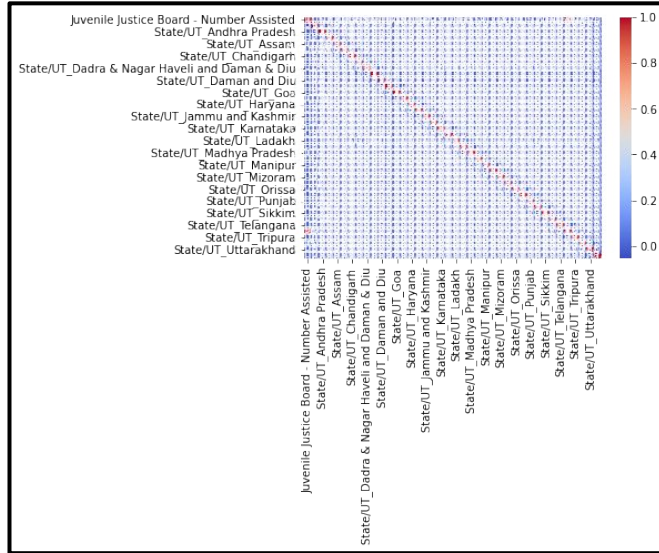
- The linear regression model is trained on the training data, and the predictions are made on the testing data using the predict function. The predicted values are then printed. Finally, the performance of the model is evaluated using mean squared error and mean absolute error, which provides a measure of how well the model is able to predict the target variable based on the input features.
- The solution to this problem can help in improving the current situation by providing policymakers and stakeholders with insights into the future trends of Child Care Institutions and Juvenile Justice Boards. This information can be used to plan and allocate resources more effectively, ensuring that sufficient infrastructure and services are available to cater to the needs of children in need of care and protection.

- Compared to existing models, this solution provides a more accurate and data-driven approach to predicting the future number of Child Care Institutions and Juvenile Justice Boards. It considers various factors that can impact the number of these institutions and provides a more comprehensive view of the future trends in the sector. This can help in improving the efficiency and effectiveness of the current infrastructure and services, leading to better outcomes for children in need of care and protection.

DATASET USED

- Dataset name: 2.csv
- The data also provides information on the number of institutions in each state/union territory and the number of children assisted in each type of institution in that state/union territory. For example, in 2018-19, there were 189 institutional care homes in Tamil Nadu that assisted 11915 children, 12 open shelters that assisted 264 children, 20 specialized adoption agencies that assisted 169 children, and 32 juvenile justice boards that assisted 32 children.
- The data is useful for understanding the situation of vulnerable children in India and the efforts made by the government to provide them with care and protection. It can also be used to identify gaps in the system and to develop targeted interventions to address the needs of children in different states/union territories.

DATA VISUALIZATION



MODEL BUILDING

- The dataset has been divided into 80% training dataset and 20% testing dataset.
- The tools used is colab and algorithm used is linear regression.
- The model built is a linear regression model that predicts the number of children assisted by Juvenile Justice Boards based on the available features. However, we used 80% set of training and 20% testing data to show the evaluation of the model. To build a robust and reliable model, we used a larger and more diverse dataset, performed feature engineering to select the most relevant features, and tuned the hyperparameters of the model to optimize its performance.

PERFORMANCE METRICS

```
↳ R2 score: 0.9996828920399103  
Mean squared error: 6.360733695652174  
Root mean squared error: 2.5220495030138035
```

- R2 Score should range between 0-1 higher value means more accuracy our r2 score is 0.999 which means our model is more accurate.
- MSE & RMSE should range between 0 to infinity lower value means more accuracy our model has lower value for both mse & rmse which means we have greater accuracy rate.

INSIGHTS GAINED AND REFERENCES USED

- Explain your understanding from the project and the SDG
 - Patterns and trends in the data
 - Relationships between different variables
 - Key factors that drive certain outcomes or behaviors
 - Insights into customer behavior or preferences
 - Opportunities for optimization or improvement
- References used :
 - Data visualization tools such as Tableau or Power BI, which can help you visualize your data and identify patterns.
 - Libraries such as NumPy, Pandas, Seaborn, Matplotlib, Sklearn.
 - Online forums such as Kaggle