Hack The Future 2025

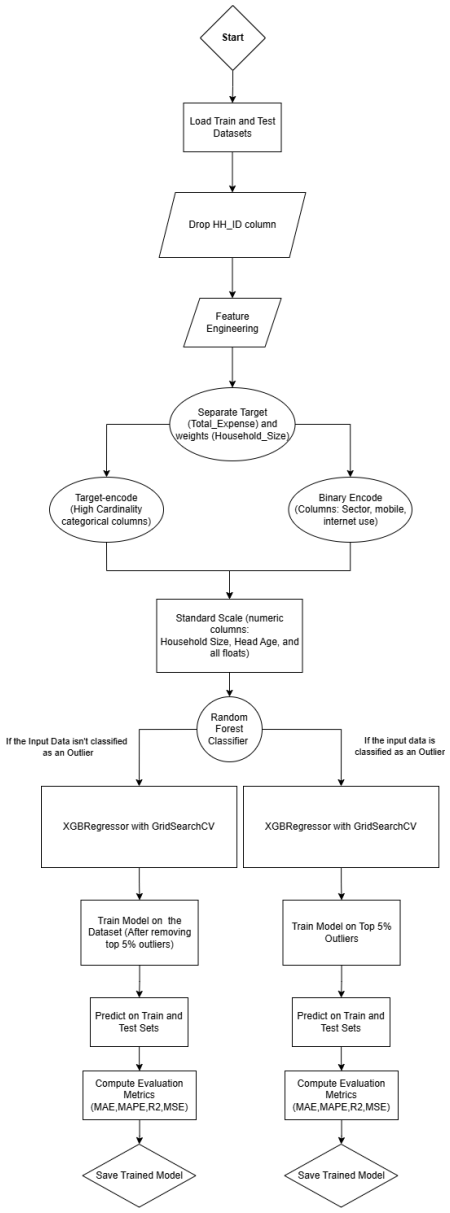
Team Nexus

Github Link: https://github.com/SayujGupta2005/Team_Nexus-HackTheFuture2025

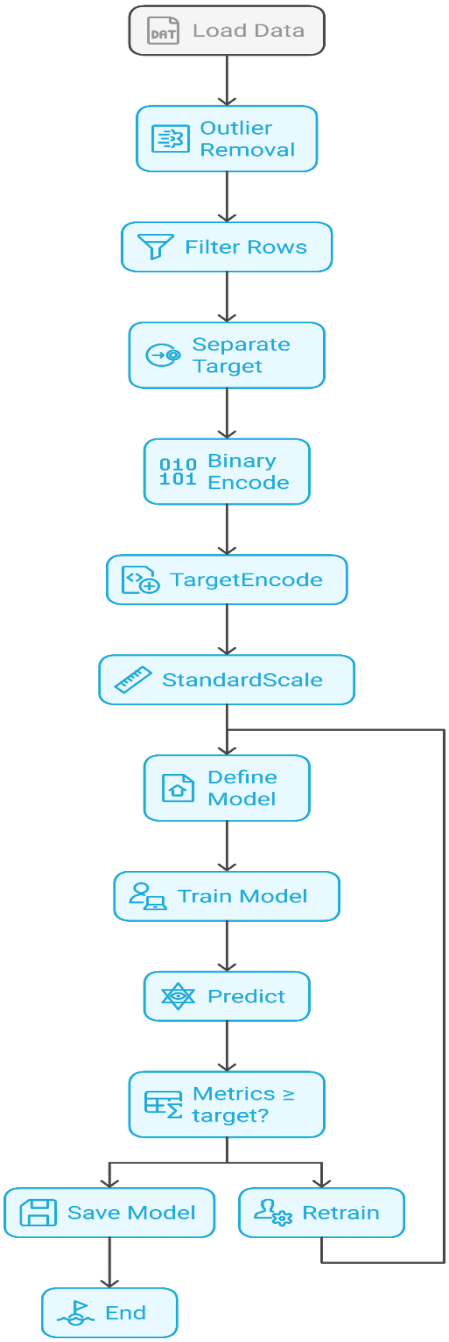
User Guide

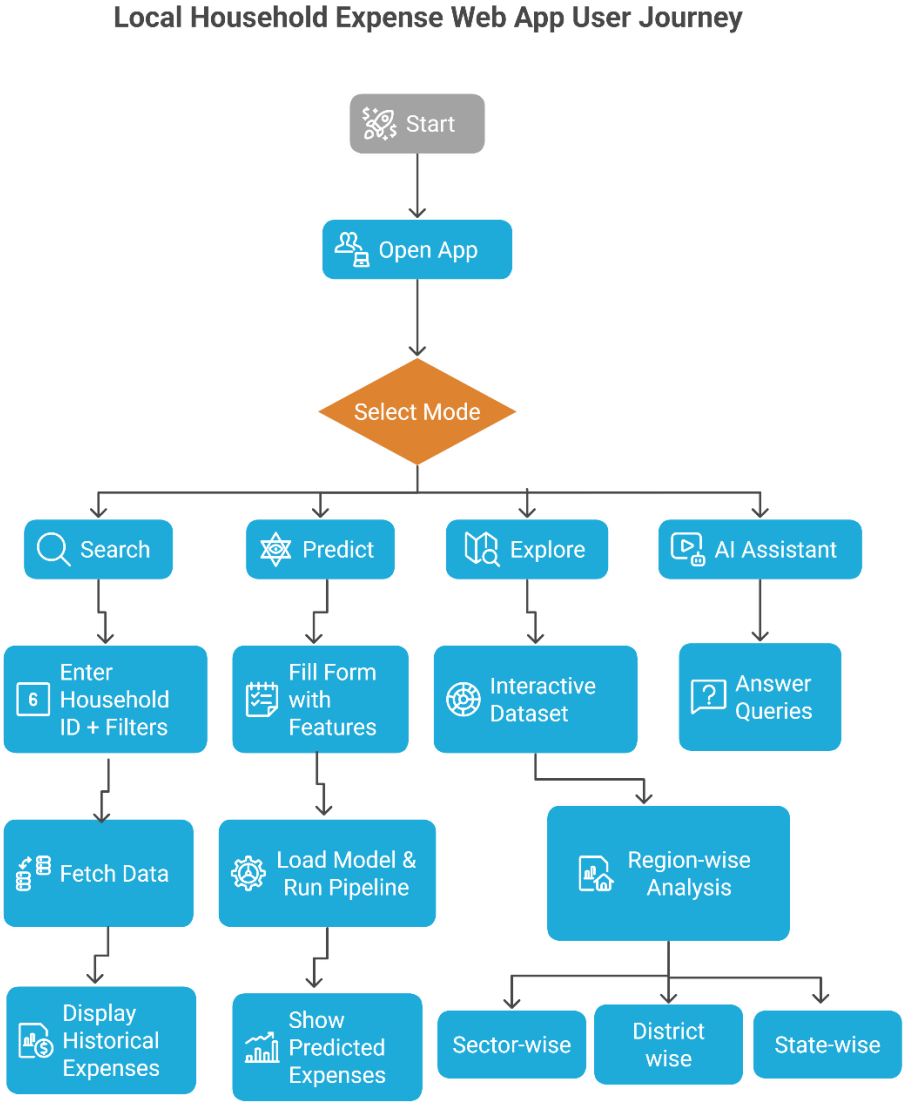
A) Hardware and Software requirements:
   - Hardware needs: Basic laptop 4gb RAM
   - Software needs: Python, VSCode, Nesstar

B) Approach:



Final XGBoost Regression Pipeline (v3)





C) List of Python libraries used:

- pandas
- numpy
- scikit-learn
- category_encoders
- xgboost
- joblib
- matplotlib
- seaborn
- flask
- CORS

D) Environmental Details:

- Google Colab

- Kaggle Notebook

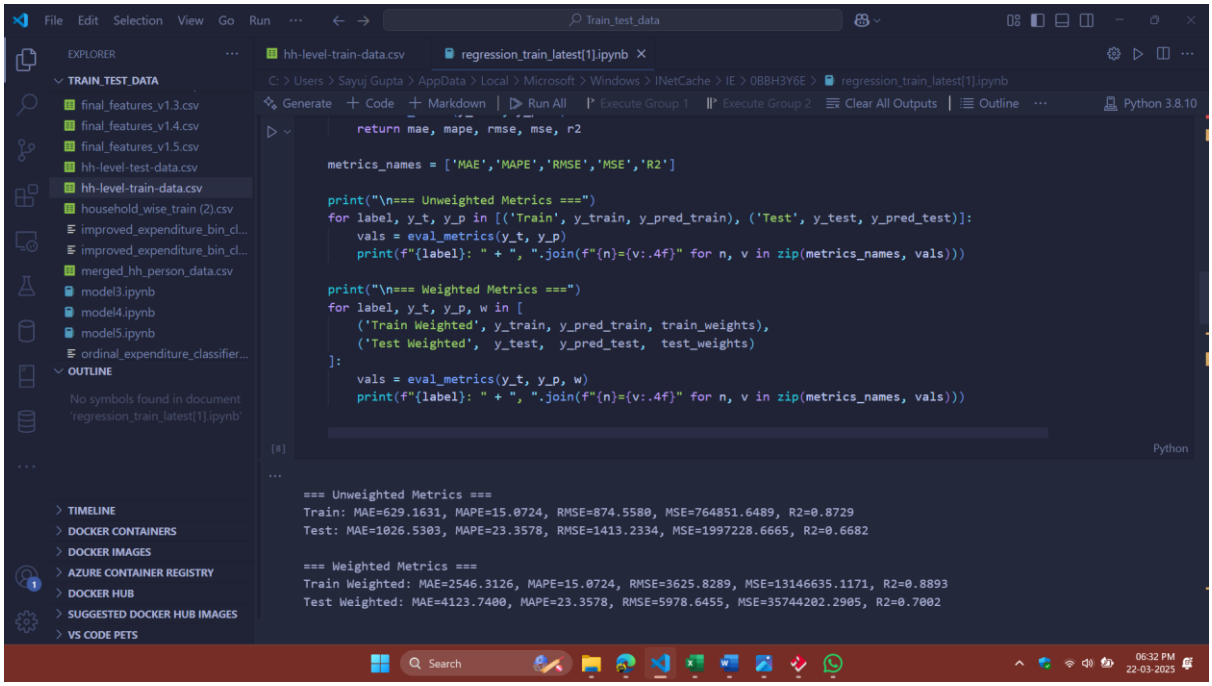
E) List of APIs Used:
- Groq – Open Source

F) List of files required to execute the program:
- For website:
  i. index.html
  ii. style.css
  iii. script.js
  iv. app.py
  v. chat.py
  vi. hh-level-train-data.csv
  vii. hh-level-test-data.csv
  viii. xgb_regressor_model_below95.pkl
  ix. xgb_regressor_model_above95.pkl
  x. rf_classification_model.pkl
- For Training:
  i. hh-level-train-data.csv
  ii. hh-level-test-data.csv
  iii. person-level-train-data.csv
  iv. person-level-test-data.csv
  v. data_preprocessing.ipynb
  vi. regression_train.ipynb
  vii. classification_train.ipynb

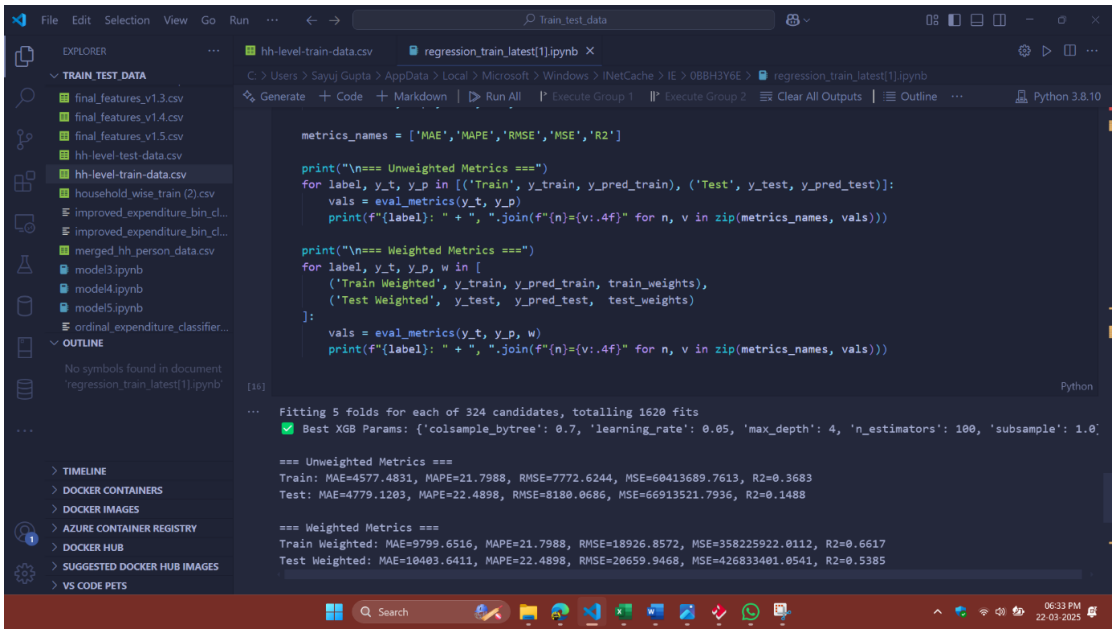
G) Steps to execute:
- For Training:
  i. Open the data_preprocessing.ipynb file.
  ii. Ensure that the train and test datasets path are defined properly.
  iii. Run the file to the end, thus creating pre-processed train and test datasets.
  iv. Open the classification_train.ipynb file.
  v. Use the pre-processed datasets to train the classifier which will classify between above and below 95%ile of Total Expenses.
  vi. Open the regression_train.ipynb file.
  vii. Use the preprocessed datasets to train both regressors.
- For website:
  i. Ensure index.html, script.js, style.css, app.py and chat.py are in same directory.
  ii. Run app.py
  iii. Open index.html in web browser.

H) Screenshots of Execution:

Regression below 95%ile:



Regression above 95%ile:



Classification:

```python
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import pickle

train_df = pd.read_csv("/home/puru/Downloads/final_train_v3.csv")
test_df = pd.read_csv("/home/puru/Downloads/final_test_v3.csv")

threshold = train_df['Total_Expense'].quantile(0.95)
train_df['is_top_5_percent'] = (train_df['Total_Expense'] >= threshold).astype(int)


X = train_df.drop(columns=['Total_Expense', 'is_top_5_percent'])
y = train_df['is_top_5_percent']

X = X.drop(columns=[col for col in X.columns if 'id' in col.lower()], errors='ignore')
test_X = test_df[X.columns]
test_df['is_top_5_percent'] = (test_df['Total_Expense'] >= threshold).astype(int)


model = RandomForestClassifier(
    n_estimators=100,
    max_depth=6,
    random_state=42
)
                                    (variable) test_preds: ndarray

model.fit(X, y)


test_preds = model.predict(test_X)


y_test = test_df['is_top_5_percent']
accuracy = accuracy_score(y_test, test_preds)
print(f"✅ Accuracy on Test Set: {accuracy:.4f}")


with open('random_forest_model2.pkl', 'wb') as f:
    pickle.dump(model, f)

✓ 16.5s                                                    Python
✅ Accuracy on Test Set: 0.9596
```

Contact Details of Team Members:

1) Sayuj Gupta
    a. Email ID: 2023ucs0112@iitjammu.ac.in
    b. Mobile No.: 9464747414
    c. Institute: Indian Institute of Technology Jammu
    d. Course: Computer Science and Engineering
    e. Year of passing: 2027
    f. Contribution:
        i. Team Leader
        ii. Feature Engineering and Model Testing
        iii. Presenter
2) HIMANSHU
    a. Email ID: 2023ucs0092@iitjammu.ac.in
    b. Mobile No.: 8595756912
    c. Institute: Indian Institute of Technology Jammu
    d. Course: Computer Science and Engineering
    e. Year of passing: 2027
    f. Contribution:
        i. Website builder

        ii. Feature Engineering

3) D Barghav
    a. Email ID: 2023ume0253@iitjammu.ac.in
    b. Mobile No.: 9790772220
    c. Institute: Indian Institute of Technology Jammu
    d. Course: Mechanical Engineering
    e. Year of passing: 2027
    f. Contribution:
        i. Dataset preparation
        ii. Model Training

4) Purushartha Gupta
    a. Email ID: 2023uce0062@iitjammu.ac.in
    b. Mobile No.: 9369573768
    c. Institute: Indian Institute of Technology Jammu
    d. Course: Civil Engineering
    e. Year of passing: 2027
    f. Contribution:
        i. Dataset preparation
        ii. Model Training

5) Parth Sachdeva
    a. Email ID: 2023ucs0105@iitjammu.ac.in
    b. Mobile No.: 9149663246
    c. Institute: Indian Institute of Technology Jammu
    d. Course: Computer Science and Engineering
    e. Year of passing: 2027
    f. Contribution:
        i. Dataset preparation
        ii. Model Training
        iii. Feature Engineering