CALIFORNIA (USA)

PREDICTION HOUSE PRICE

DSAI

MEMBERS

KHAMTAN KHAMSAMSEE

685380020-3

SEARCH

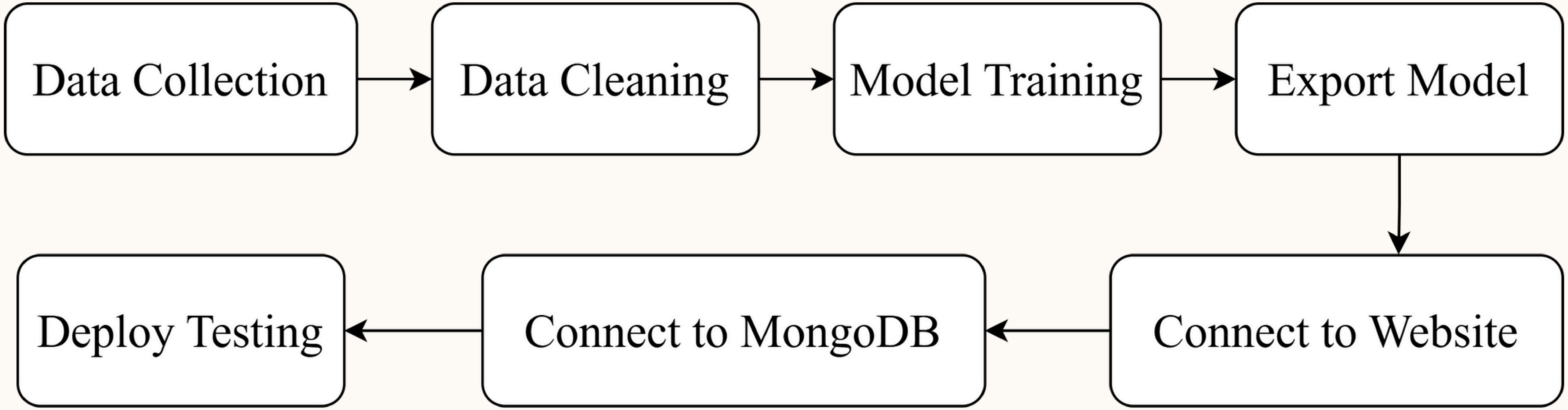
ANAPAT CHANSONG

685380035-0

POONYAWAT MANDEE

685380025-3

PIPELINE BACKEND



DATA

The dataset for this model

- **Collection : Kaggle datasets**
  - The House Price data comes from Kaggle datasets.
  - https://cphaigh.github.io/KaggleProjects/kaggleHousePrices/HousePrices12_19.html?utm_source=chatgpt.com#libraries

How to clean ?

The columns which contains a large number of NaN values (Alley, PoolQC, and MiscFeature) were dropped from CSV Datasets. Missing numeric values were imputed using the mean of each column.

9 key features selected: (OverallQual, TotalBsmtSF, LotArea, GarageCars, Fireplaces, BedroomAbvGr, GrLivArea, FullBath, Neighborhood)

MACHINE LEARNING MODEL

•RANDOM FOREST REGRESSOR

- “We predict house prices with a Random Forest. Think of it as many decision trees voting together.
- Each tree sees a slightly different sample and a different subset of features, then we average the results.”
- “That teamwork cuts variance and lets the model catch non-linear patterns and feature interactions
- for example, how living area and neighborhood combine to drive price.”
- “It needs little feature scaling, is fairly robust to outliers, and works well on real-world housing data.”

RESULT

- **RANDOM FOREST REGRESSOR**

Accuracy k-fold CV:

$R^2 = 0.956$, MAE = $9,703.72

Exporting Model as joblib.

MACHINE LEARNING MODEL

•KNN IMPUTER

- “Before predicting, we fill missing numeric fields using K-Nearest Neighbors.
- We look for homes most similar to the one being entered and borrow their values to fill the gaps.”
- “This is more realistic than just using a global mean because it respects local context—similar homes tend to have similar numbers.”
- “For categorical fields, we use the most frequent value or an explicit ‘None’ when the feature truly doesn’t exist.”

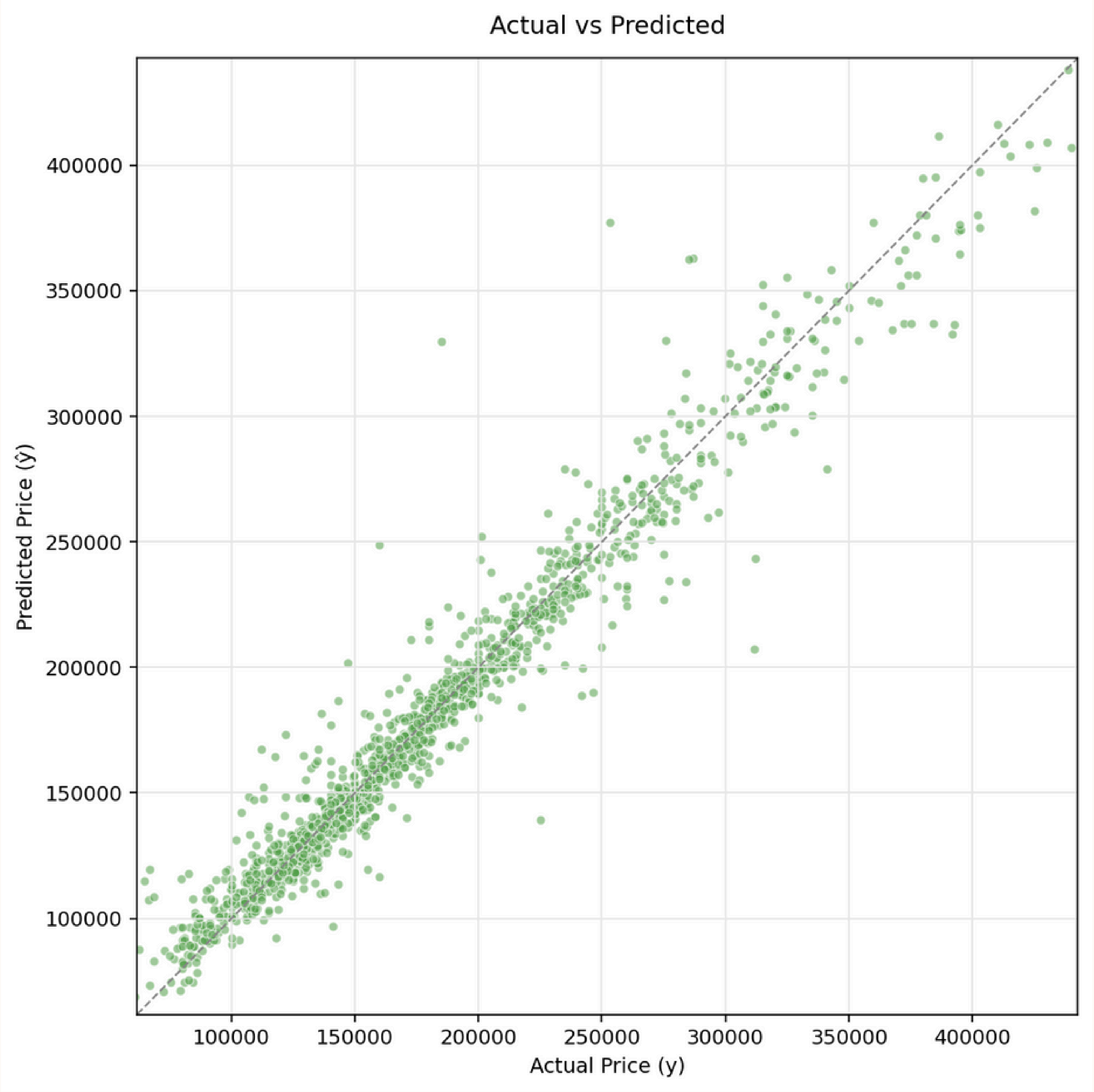
RESULT

•KNN IMPUTER

```
--- Imputer Performance Metrics ---
Feature: 'GarageCars'
   MAE (Mean Absolute Error): 0.3836
   RMSE (Root Mean Squared Error): 0.6621

Feature: 'TotalBsmtSF'
   MAE (Mean Absolute Error): 237.9986
   RMSE (Root Mean Squared Error): 335.7427
```

SUMMARY STRONG LINEAR RELATIONSHIP BETWEEN ACTUAL AND PREDICTED



•RANDOM FOREST REGRESSOR

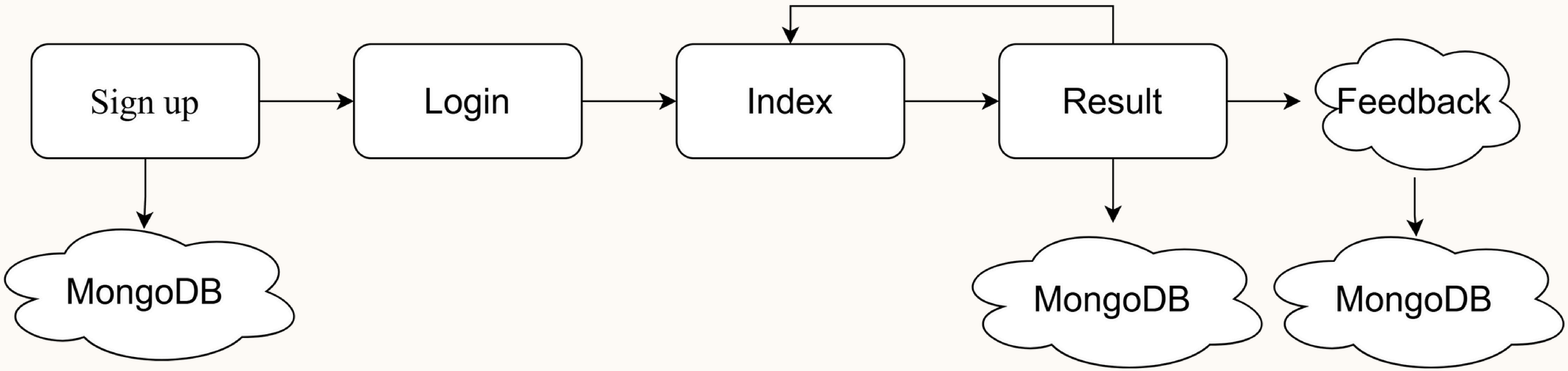
Summary Strong linear relationship with most points close to the (y=x) line → good overall fit (high (R^2)). As actual prices rise, the spread widens (heteroscedasticity), and many high-price cases fall below the line → underprediction at the top end. Conclusion: strong performance in low–mid ranges; increasing error and slight low bias for expensive properties.

MACHINE LEARNING MODEL

EXPORTING MODEL

After training model, the model is exported to the joblib file. Using joblib is more suitable because it is faster, more memory-efficient, and specifically optimized for machine learning models that rely on NumPy and scikit-learn. It ensures smooth loading in the Flask/FastAPI backend and aligns with best practices for model deployment.

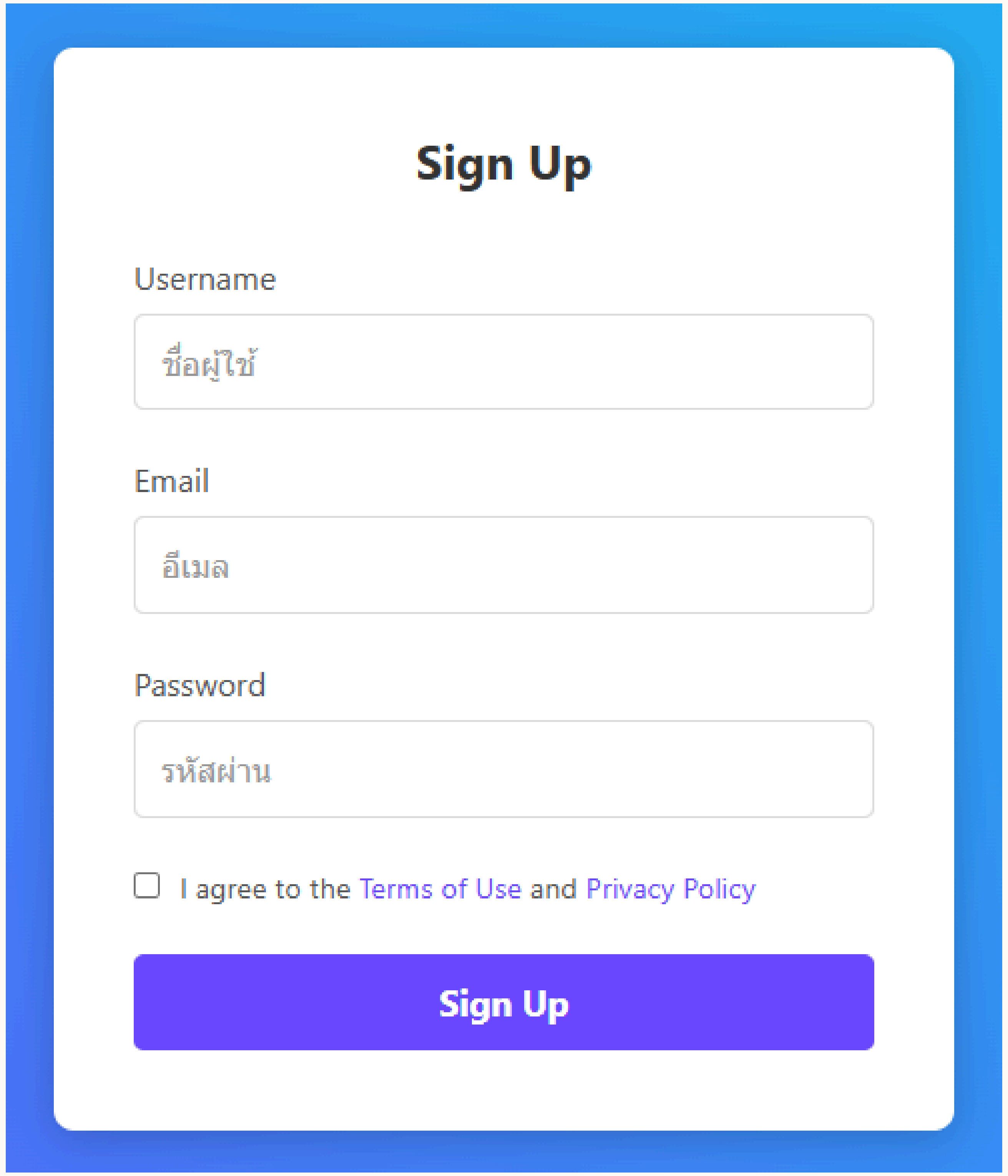
PIPELINE FRONTEND

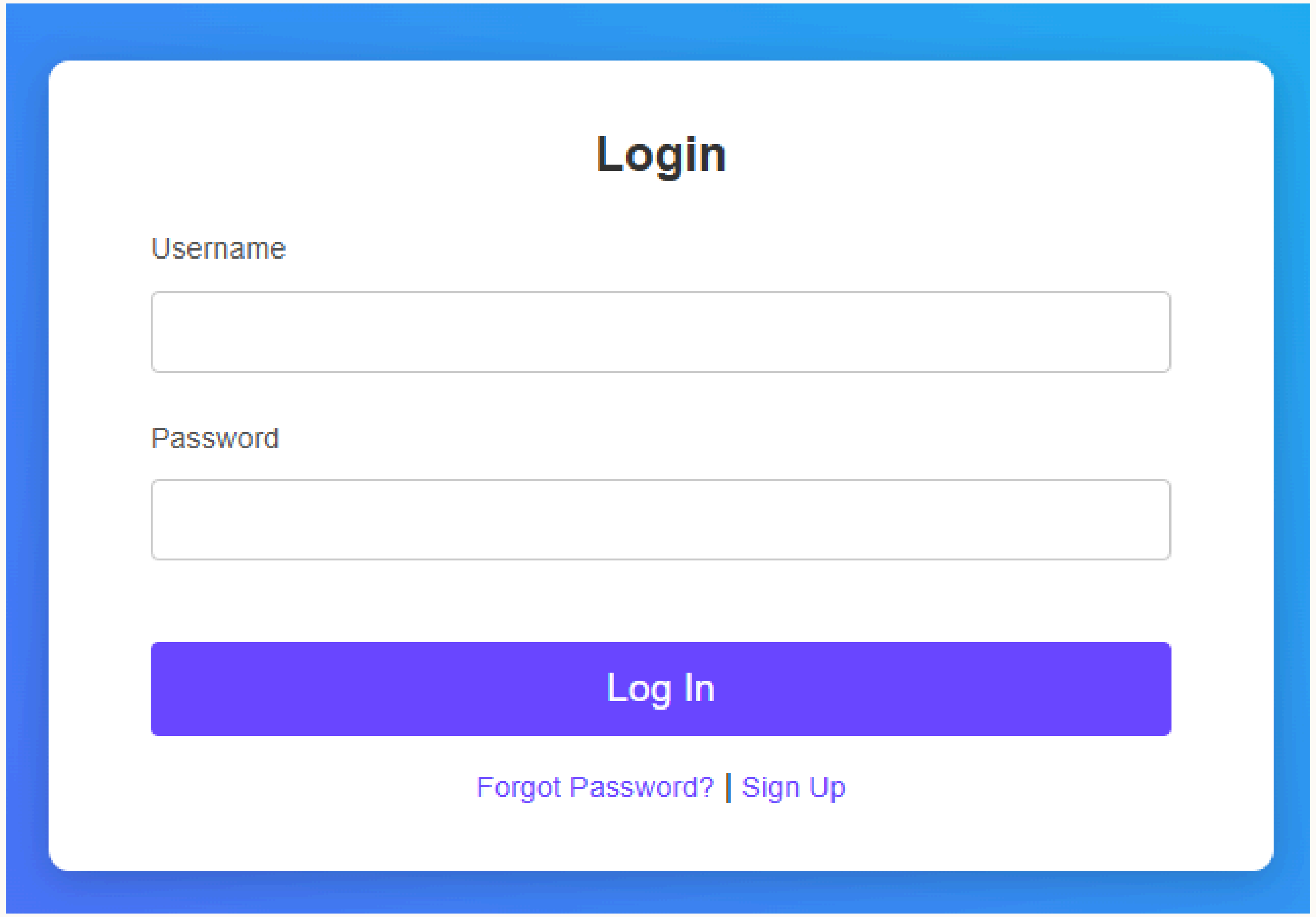


WEBSITE IMPLEMENTATION

- Sign up and Login
- Privacy Policy and Terms of Use
- Index
- Result

SIGN UP AND LOGIN



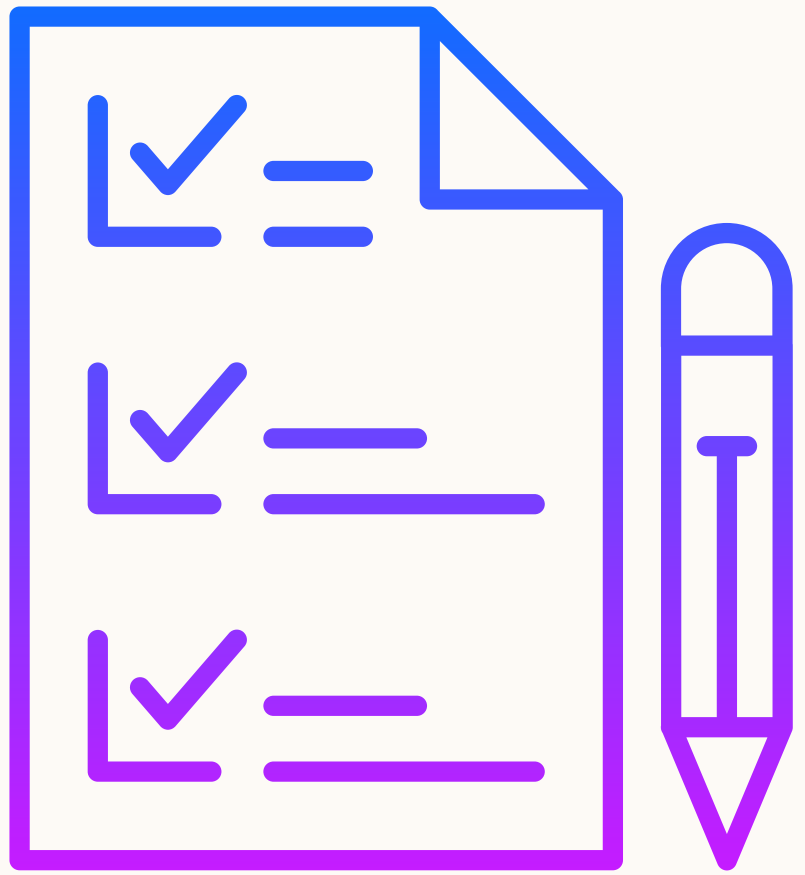


PRIVACY POLICY

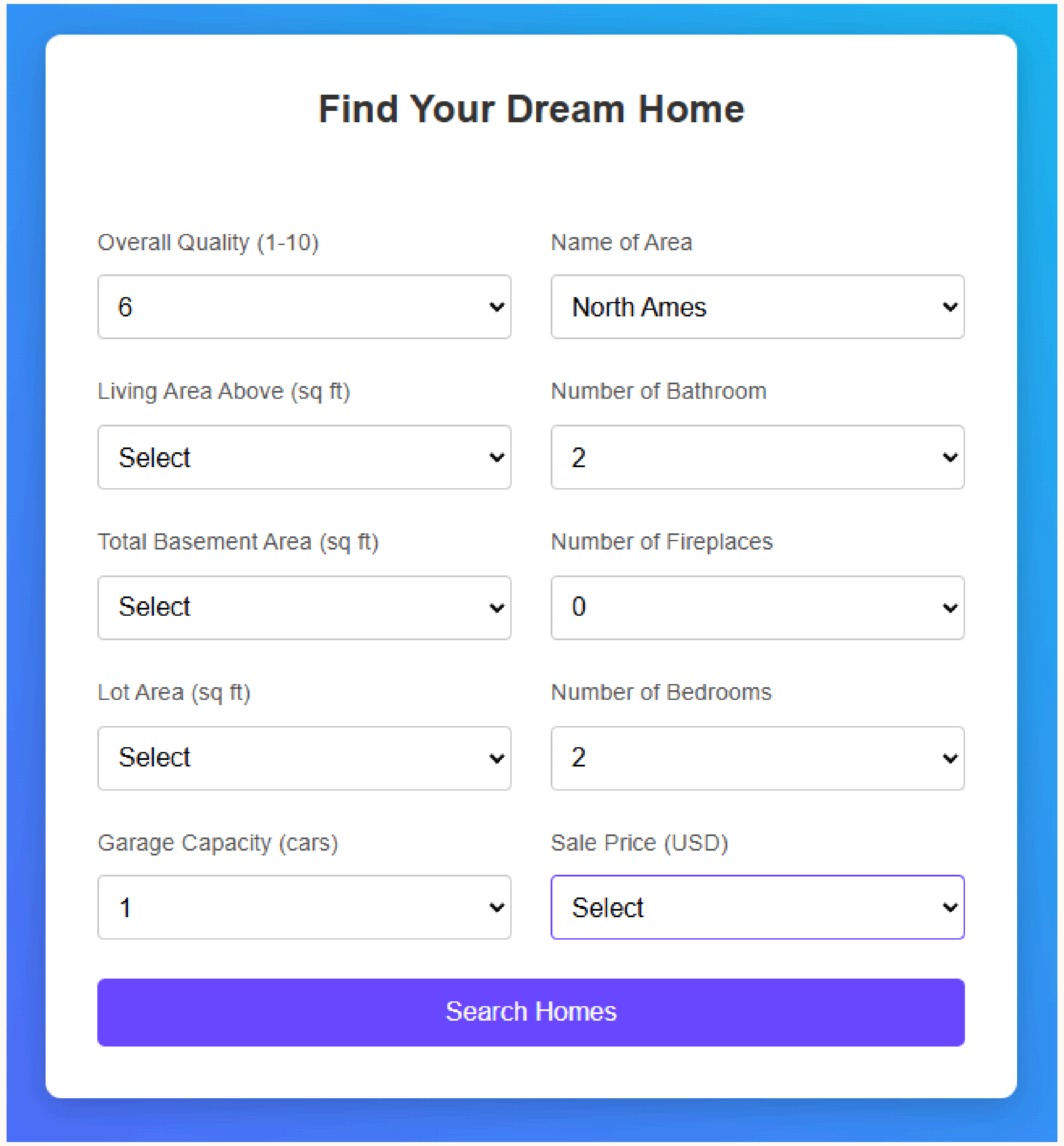
- Data Collection: Collects only necessary data (name, email, password, search history, feedback) for user verification, predictions, and app improvement.
- Consent: Users provide explicit consent via in-app notifications before data collection.
- Security: Data is encrypted, and access is restricted to authorized purposes. The full policy is accessible on the website.
- https://cppa.ca.gov/regulations/

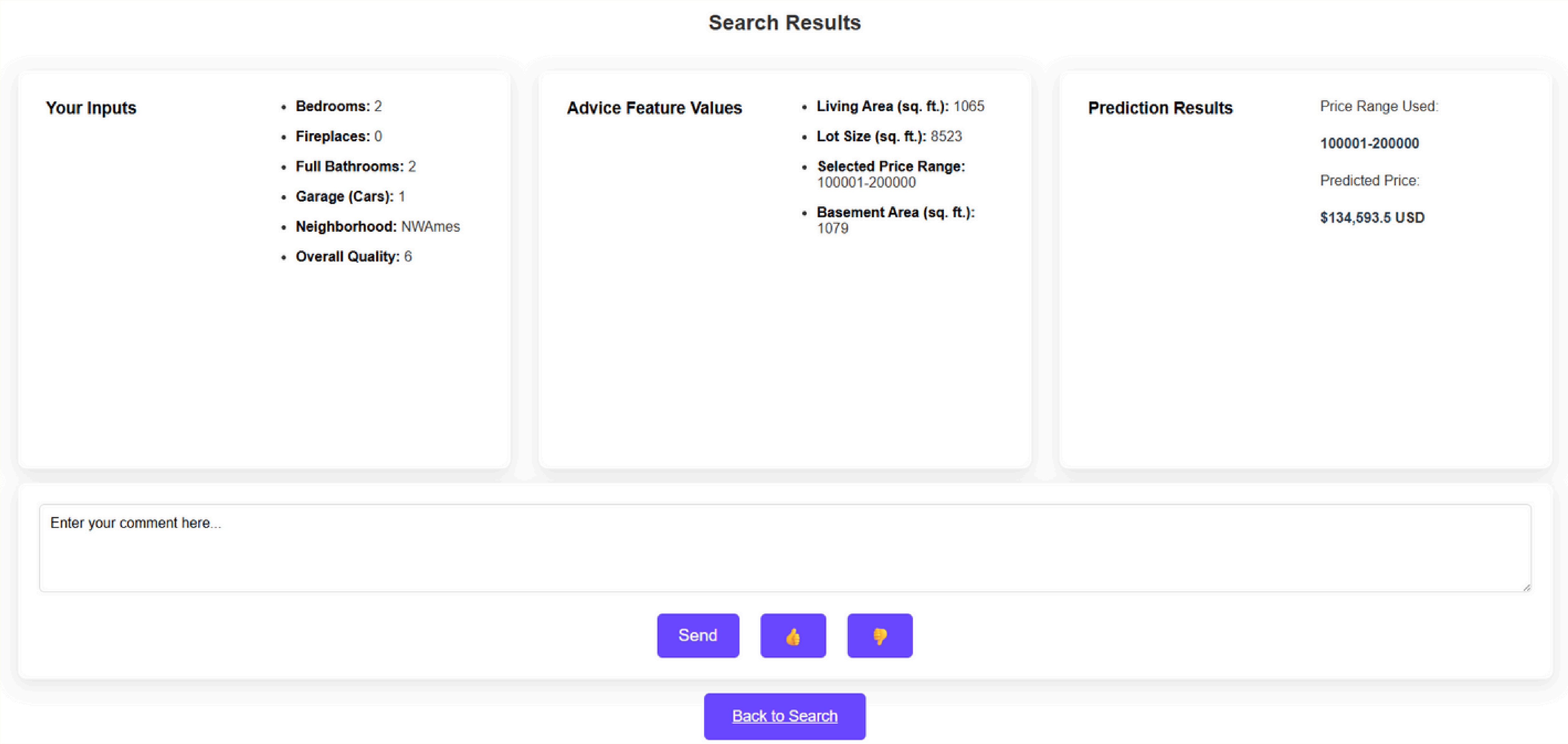
TERM OF USE

- Acceptable Use: Users must use predictions for personal or professional decision-making, not for unlawful purposes.
- Liability: Predictions are estimates; the system is not liable for financial decisions based on outputs.
- Compliance: Aligns with CCPA/CPRA and FHA. The full terms are available on the website.



INDEX AND RESULT





USER’S DATA

User data (username, hashed password, email, search history, results, feedback and timestamp) is stored in MongoDB Atlas with user consent for model and app improvements.

HIRE ME

- Flask
- flask-cors
- bcrypt
- pymongo
- joblib
- pandas
- numpy
- scikit-learn
- fastapi
- uvicorn
- gunicorn
- flask-bcrypt


LET'S TRY
DEMO


THANK YOU