

# PROJECT DOCUMENTATION & SUBMISSION – PREDICTING HOUSE PRICES USING MACHINE LEARNING

---

## PROJECT DOCUMENTATION

**Project Title:** Predicting house prices using machine learning

### Project Steps:

- Data collection and preparation
- Feature engineering
- Model selection and training
- Model evaluation
- Model deployment

### Program:

```
#import python packages

import numpy as np

import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestRegressor


# Load the dataset

df = pd.read_csv('house_price_prediction_dataset.csv')


# Clean and prepare the data

def clean_data(df):
```

```
# Remove outliers

df = df[df['SalePrice'] < 1e7]


# Fill in missing values

df['LotFrontage'] = df['LotFrontage'].fillna(df['LotFrontage'].mean())


# Create new features

df['SquareFootagePerBedroom'] = df['GrLivArea'] / df['Bedrooms']


return df


df = clean_data(df)


# Split the data into training and test sets

X = df.drop(['SalePrice'], axis=1)
y = df['SalePrice']


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25)


# Train a machine learning model

model = RandomForestRegressor()
model.fit(X_train, y_train)


# Evaluate the model's performance

y_pred = model.predict(X_test)


print('Model accuracy:', model.score(X_test, y_test))
```

## Output:

---

Model accuracy: 0.95

---