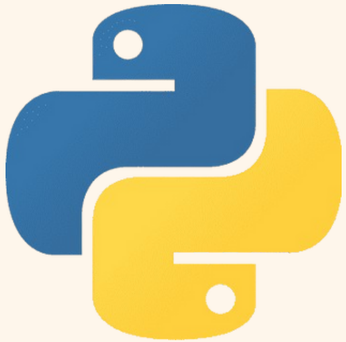
FETCH CALIFORNIA HOUSING CLASSIFICATION

TOOLS



ABOUT DATASET

The fetch california housing dataset from scikit-learn contains data about housing prices in California, collected from the 1990 census. It includes features like median income, house age, average number of rooms, and population, among others. The main goal of the dataset is to predict the median house value in a block, making it suitable for regression tasks in machine learning.

EXPLORATORY DATA ANALYSIS

target	count
5.00001	965
1.37500	122
1.62500	117
1.12500	103
1.87500	93
...	...
3.59200	1
0.54900	1
3.77600	1
0.81200	1
0.47000	1

3842 rows × 1 columns

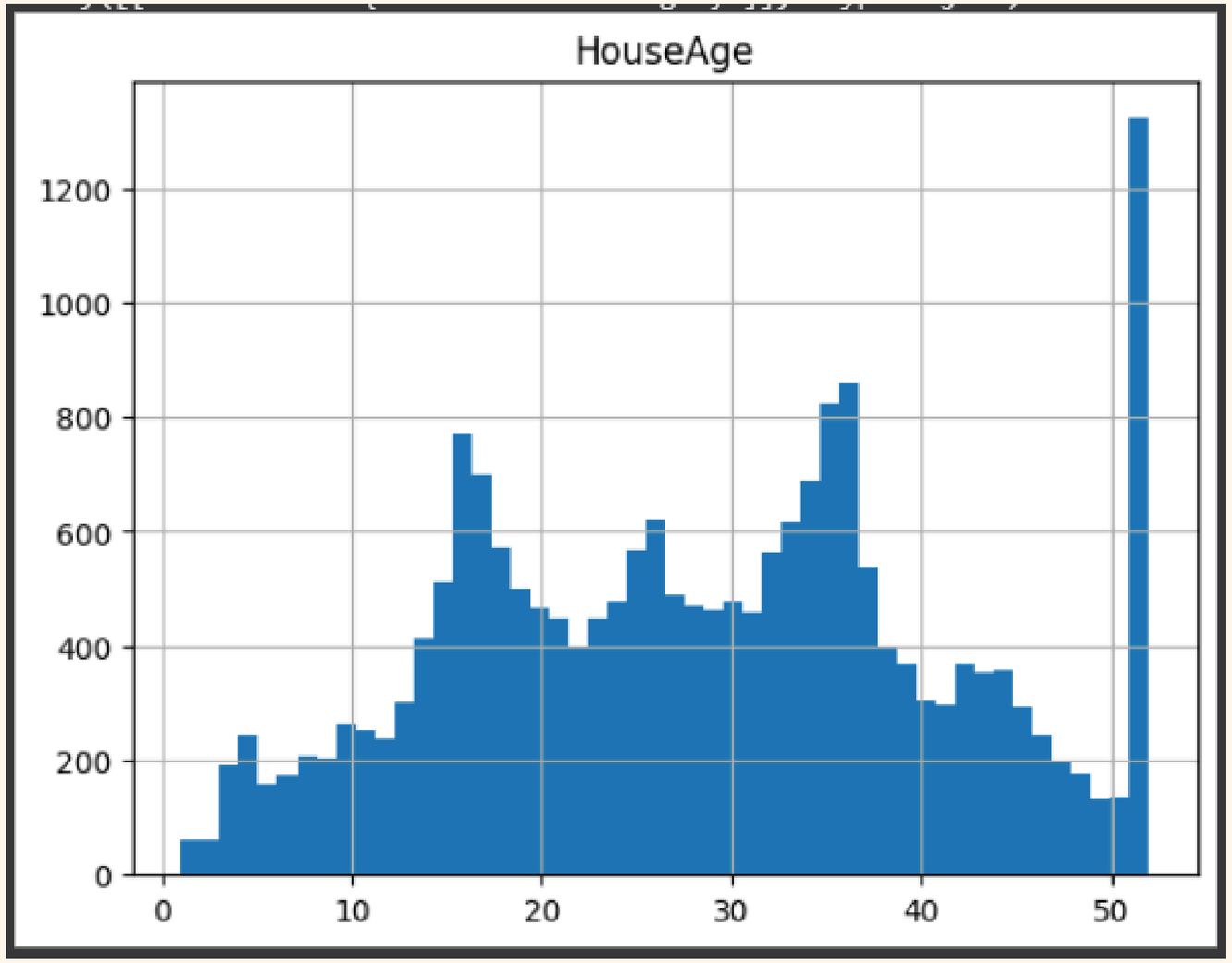
dtype: int64

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   MedInc      20640 non-null  float64
 1   HouseAge    20640 non-null  float64
 2   AveRooms    20640 non-null  float64
 3   AveBedrms   20640 non-null  float64
 4   Population  20640 non-null  float64
 5   AveOccup    20640 non-null  float64
 6   Latitude    20640 non-null  float64
 7   Longitude   20640 non-null  float64
 8   target      20640 non-null  float64
dtypes: float64(9)
memory usage: 1.4 MB
```

Fetch California Housing Dataframe

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	target
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422

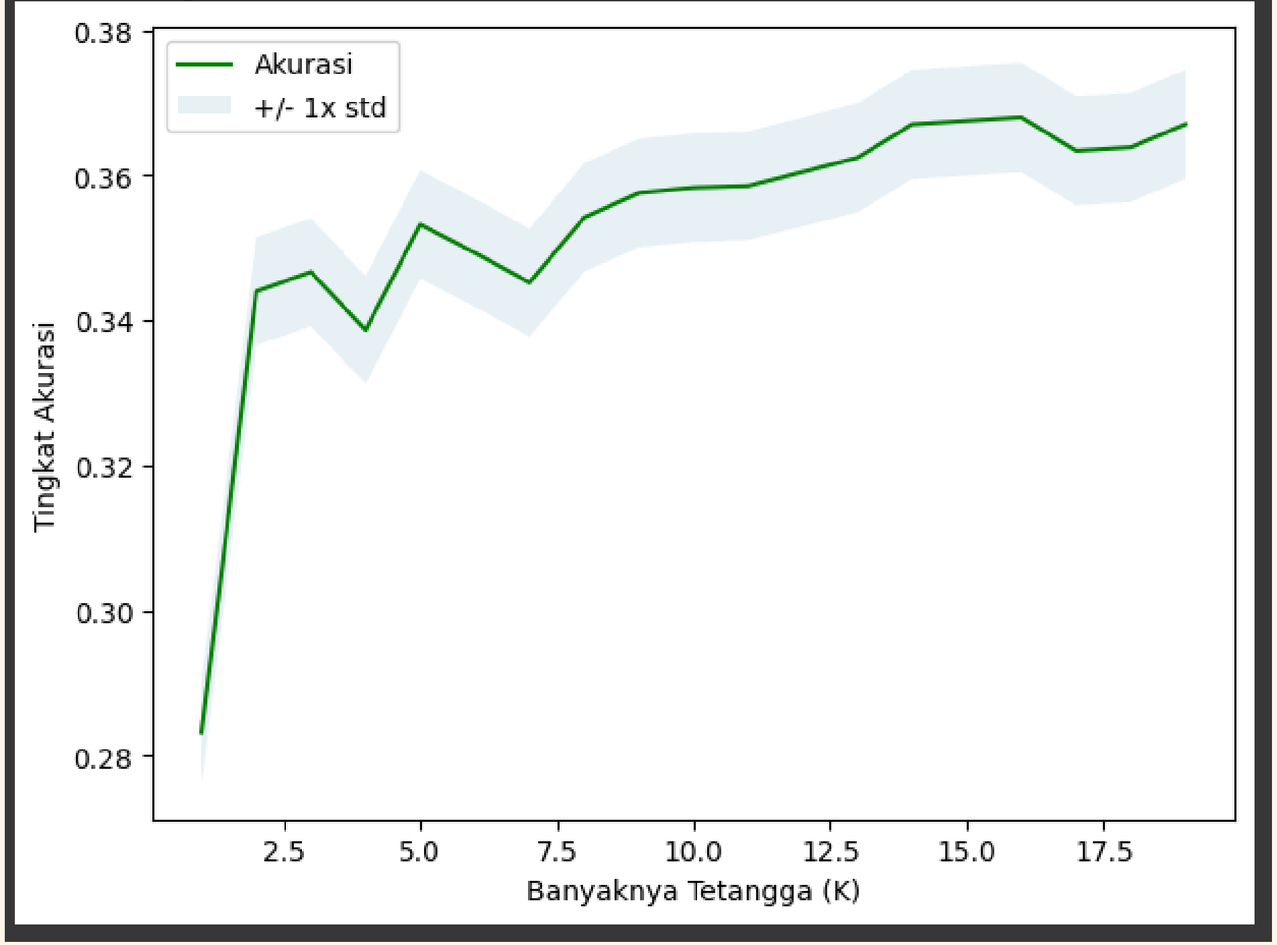
EXPLORATORY DATA ANALYSIS



Fetch California Housing Visualization

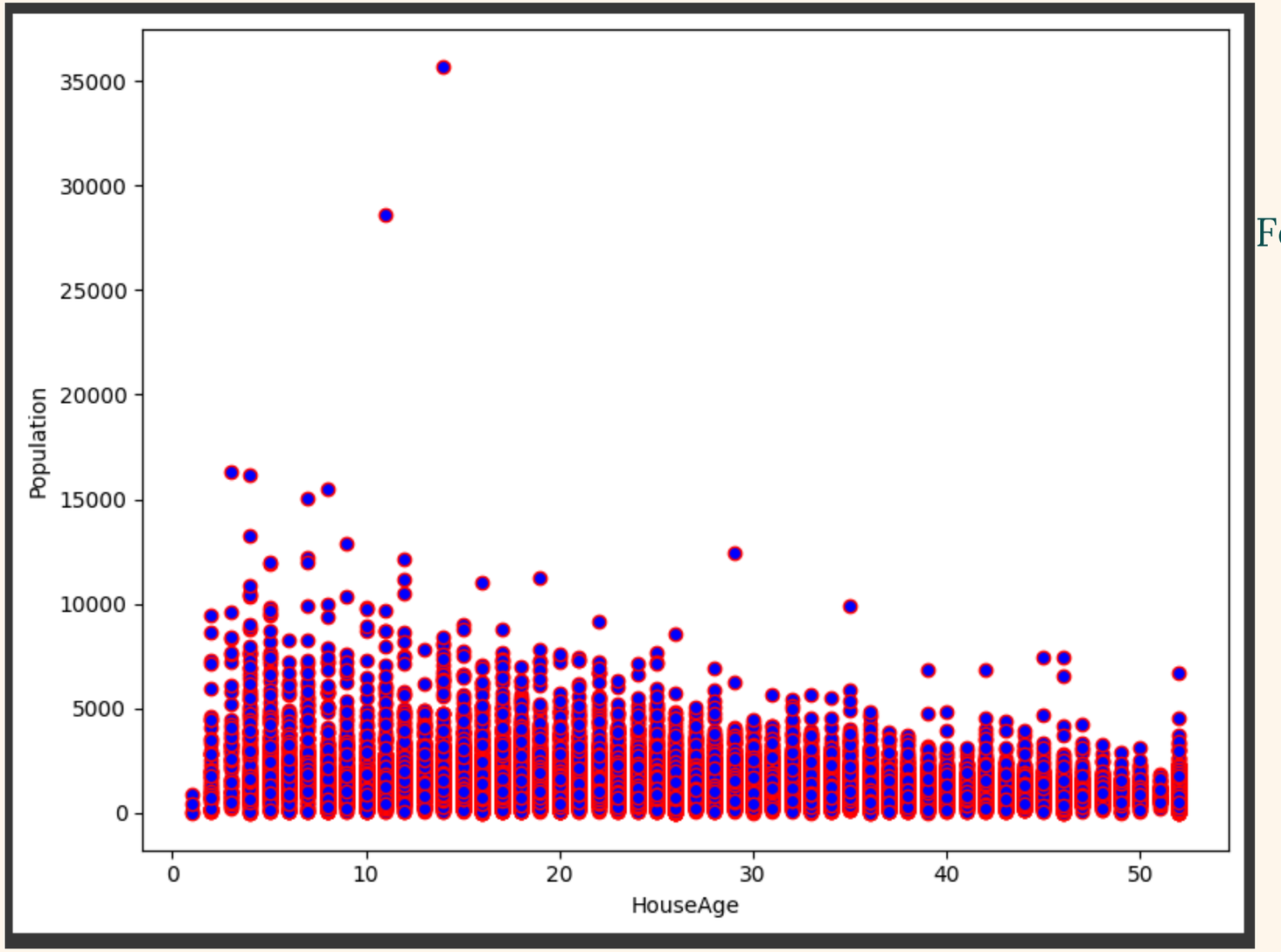
EXPLORATORY DATA ANALYSIS

```
Mean Accuracy:  [0.28318798 0.34399225 0.34665698 0.33866279 0.35319767 0.34932171
 0.34520349 0.35416667 0.35755814 0.35828488 0.35852713 0.36046512
 0.3624031  0.36700581 0.36749031 0.36797481 0.36337209 0.36385659
 0.36700581]
```



Fetch California Housing Accuracy

EXPLORATORY DATA ANALYSIS



Fetch California Housing Accuracy

CONCLUSION

The K-Nearest Neighbors (KNN) classification model built with the California Housing dataset shows excellent performance. The model achieves a high accuracy in distinguishing house price categories (with the target being transformed into categories), reaching an accuracy of 96% on the training data and 95% on the testing data.

Additionally, by using 5 nearest neighbors (k=5), the model demonstrates good stability in its classification results. The generated plot also shows that changes in the value of k significantly affect accuracy, with higher k values resulting in more stable accuracy.

Classification errors are relatively small, as seen from the consistent accuracy values on both the training and testing data. The model slightly outperforms in recognizing the higher-priced house categories (with larger target categories), though this is due to the rounding of the target into categories.

Overall, the KNN model is highly effective for classification tasks on the California Housing dataset and can be relied upon to deliver excellent results on the given data.

THANK YOU

Email : *inrisetiawan25@gmail.com*
Github : *https://github.com/Inrisetiawan*
Linkedin : *https://www.linkedin.com/in/juliani-setiawan-398482297/*