House Pricing Forecast

Zirui He
Luke McEvoy
Yinglu Wang
Dec 7th, 2021

Goal and Data Source

- Goal: predicting trends for house pricing of cities in india using different parameters
- Data Owner: Abhinav Jhanwar
- Original Data Row Number: 29k
- Memory Usage: 2.7 MB
- Columns: UNDER_CONSTRUCTION, RERA, BHK_NP., BHK_OR_RK, SQUARE_FT, READY_TO_MOVE, RESALE, ADDRESS

How we processed the dataset

- Use different packages such as numpy, pandas, seaborn, etc
- Extract all the parameters and split the parameters by comma
- House addresses contains city names; 3 tiers of cities
- Cities: bigger -> smaller population
- Classification and regression models: EDA(boxplot), KNeighborsRegressor, XGB, Decision Tree Regressor, MLPRegressor, Linear Regressor, Random Forest Regressor

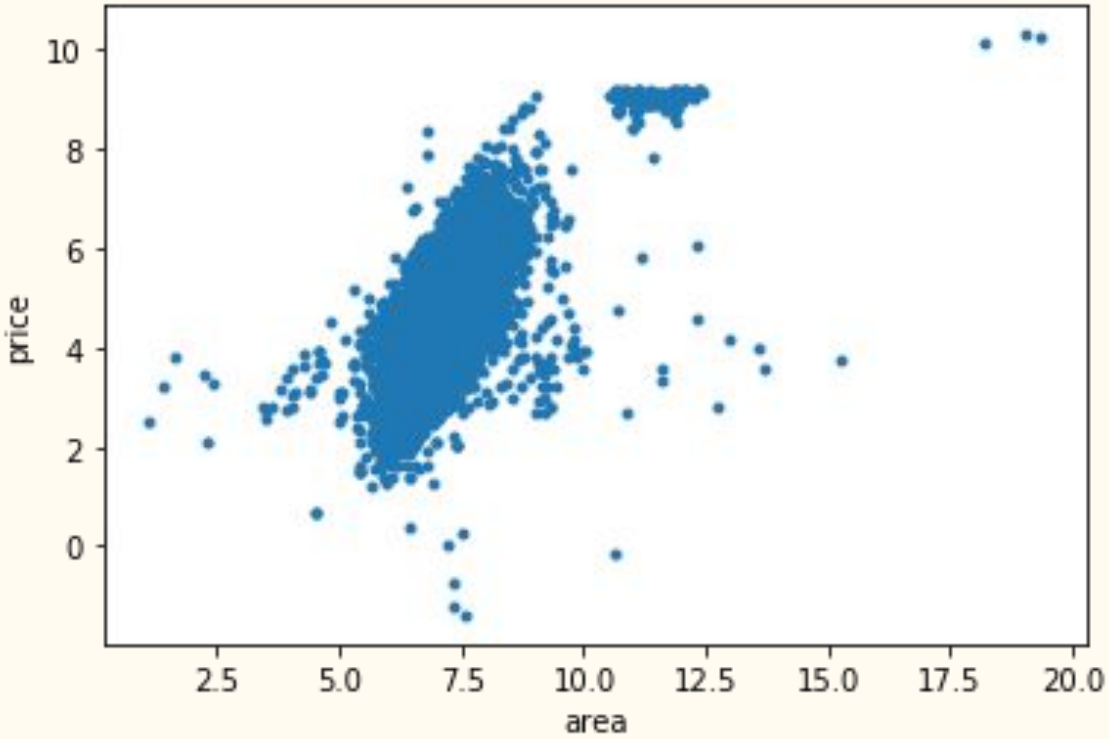
Overview of Each Column

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	READY_TO_MOVE	RESALE	ADDRESS	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)
0	Owner	0	0	2	BHK	1300.236407	1	1	Ksfc Layout,Bangalore	12.969910	77.597960	55.0
1	Dealer	0	0	2	BHK	1275.000000	1	1	Vishweshwara Nagar,Mysore	12.274538	76.644605	51.0
2	Owner	0	0	2	BHK	933.159722	1	1	Jigani,Bangalore	12.778033	77.632191	43.0
3	Owner	0	1	2	BHK	929.921143	1	1	Sector-1 Vaishali,Ghaziabad	28.642300	77.344500	62.5
4	Dealer	1	0	2	BHK	999.009247	0	1	New Town,Kolkata	22.592200	88.484911	60.5
5	Owner	0	0	3	BHK	1250.000000	1	1	South Chittoor,Kochi	10.033280	76.282571	42.0
6	Dealer	0	0	3	BHK	1495.053957	1	1	Sodala,Jaipur	26.916347	75.795600	66.5
7	Owner	0	1	3	BHK	1181.012946	1	1	Kharar,Mohali	30.740000	76.650000	52.0
8	Dealer	0	1	2	BHK	1040.000000	1	1	Bileshivale,Bangalore	13.054202	77.674002	41.6
9	Owner	0	1	2	BHK	879.120879	1	1	Chromepet,Chennai	12.951610	80.140970	36.0

Overview of Entries

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29451 entries, 0 to 29450
Data columns (total 12 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   POSTED_BY               29451 non-null  object
 1   UNDER_CONSTRUCTION      29451 non-null  int64
 2   RERA                    29451 non-null  int64
 3   BHK_NO.                 29451 non-null  int64
 4   BHK_OR_RK               29451 non-null  object
 5   SQUARE_FT               29451 non-null  float64
 6   READY_TO_MOVE           29451 non-null  int64
 7   RESALE                  29451 non-null  int64
 8   ADDRESS                 29451 non-null  object
 9   LONGITUDE               29451 non-null  float64
 10  LATITUDE                29451 non-null  float64
 11  TARGET(PRICE_IN_LACS)   29451 non-null  float64
dtypes: float64(4), int64(5), object(3)
memory usage: 2.7+ MB
```

EDA

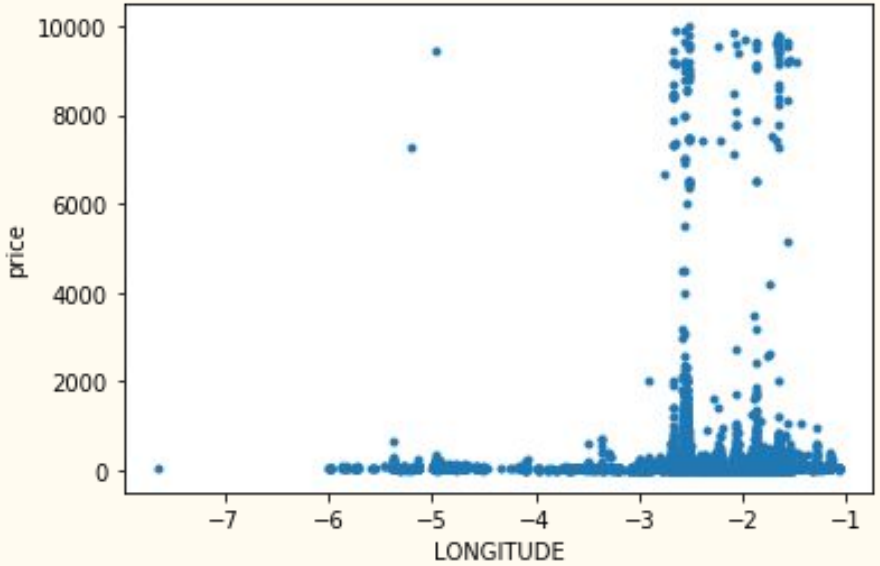


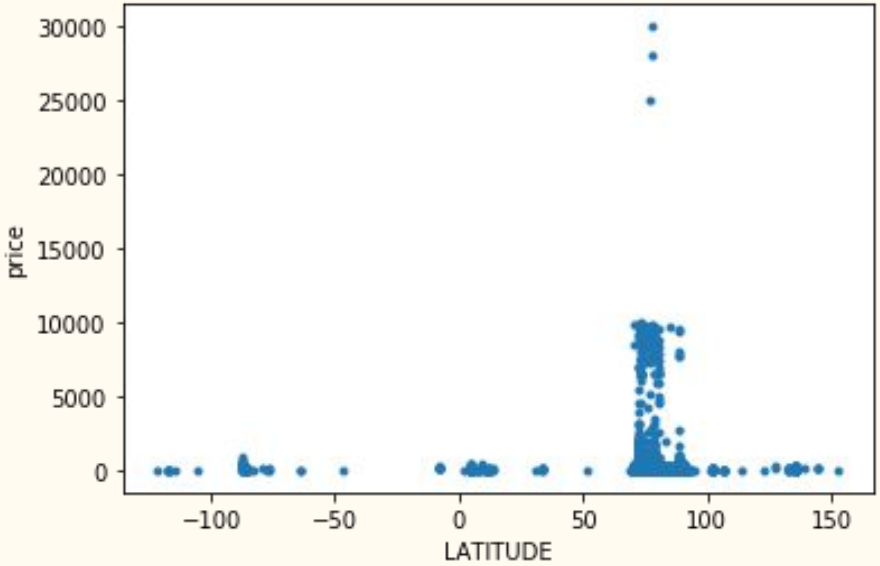
Distribution boxplot

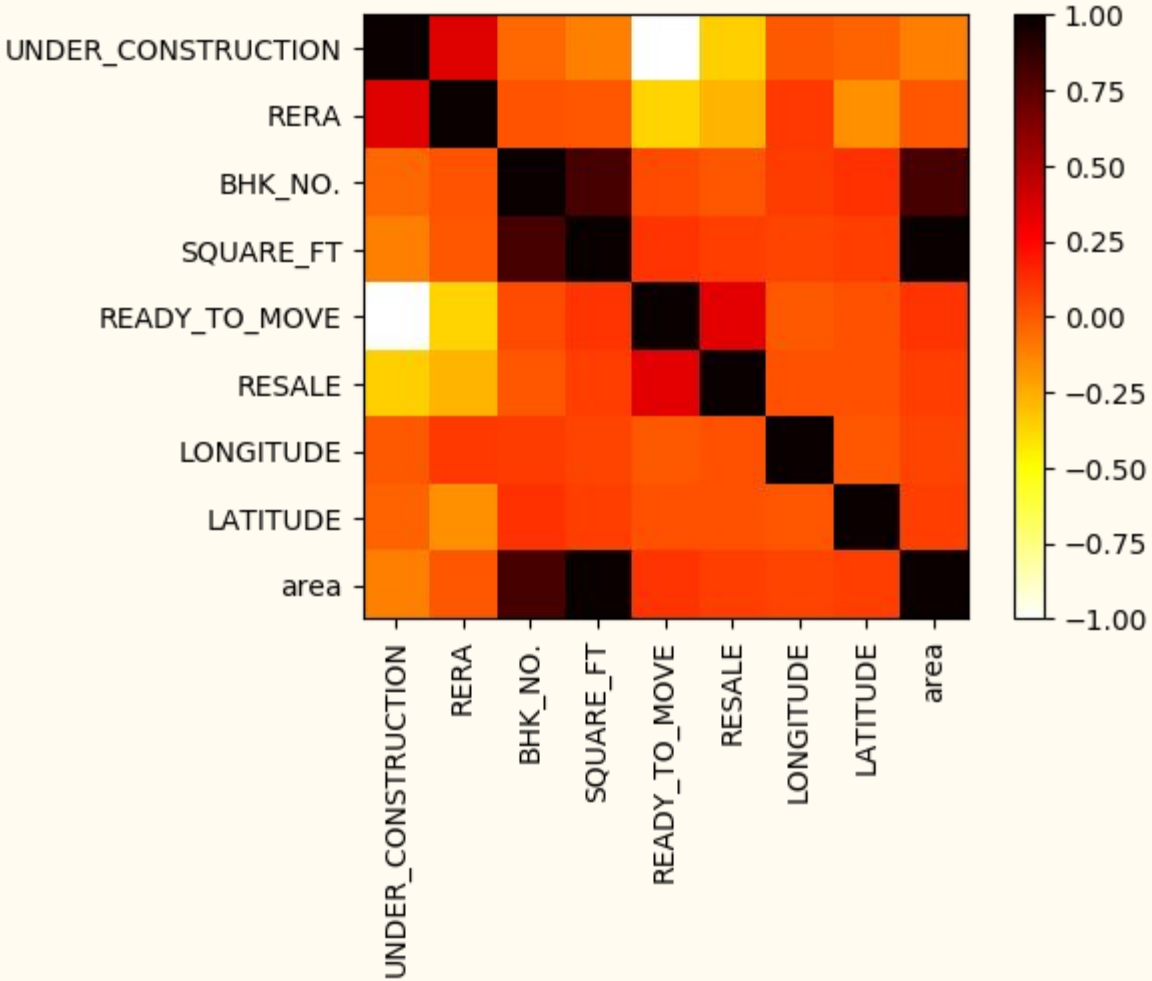
for **price** vs **area**

EDA

Distribution box plot for **price** vs **longitude** and **latitude**

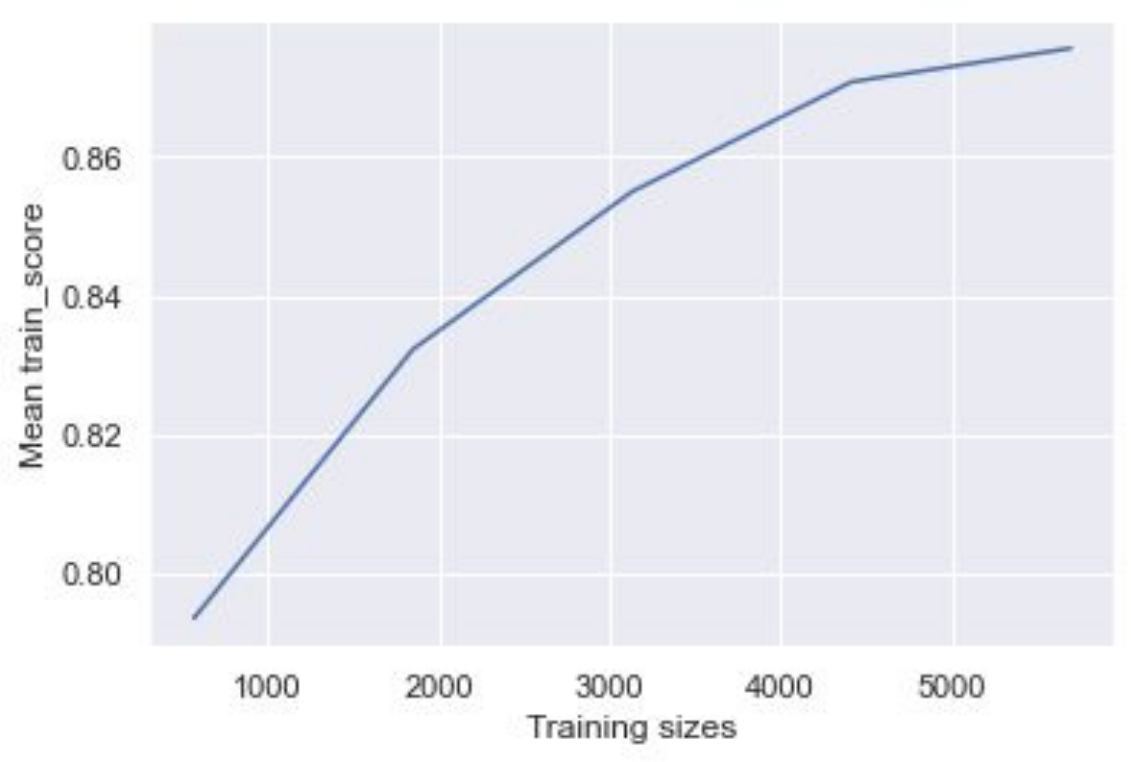






Heatmap using PLT

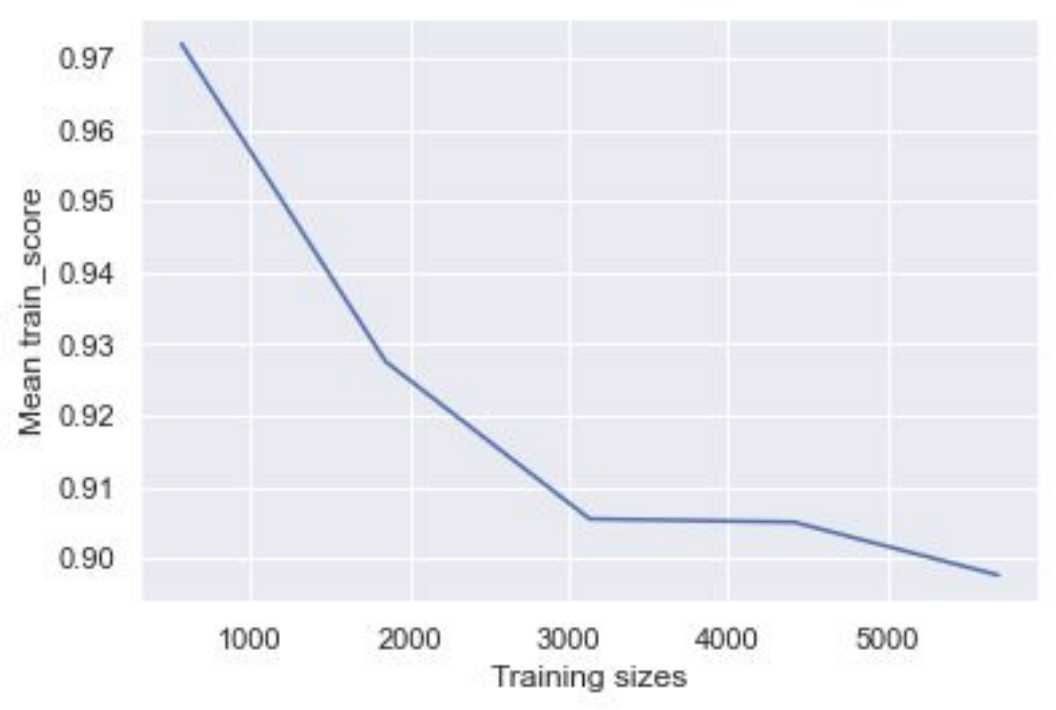
KNN Training Model



Accuracy:

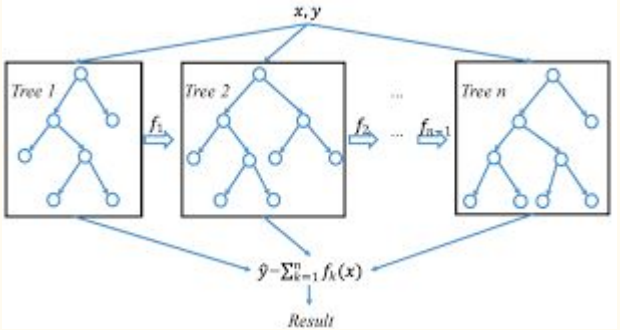
0.7937267071315093

XGBoost Training Model

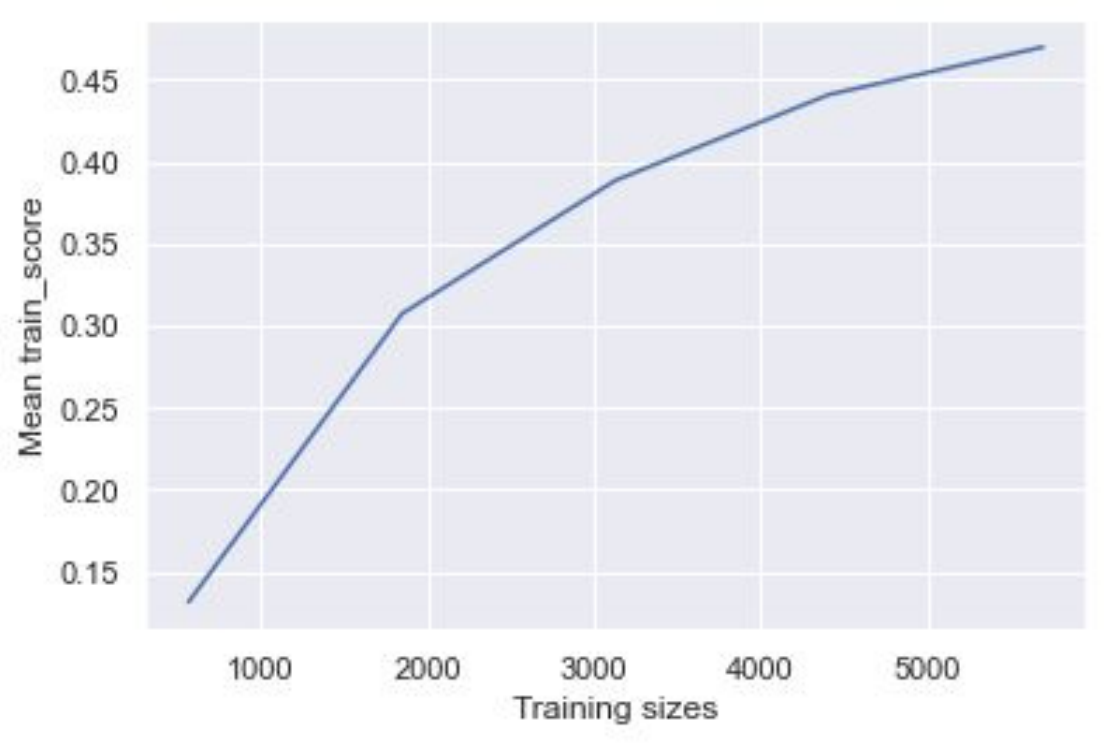


Accuracy:

0.8425897270434302

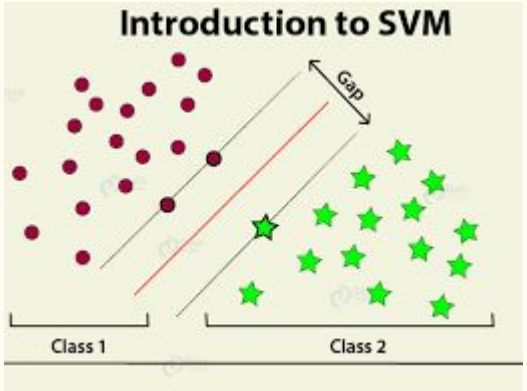


Support Vector Machine (SVM) Training Model

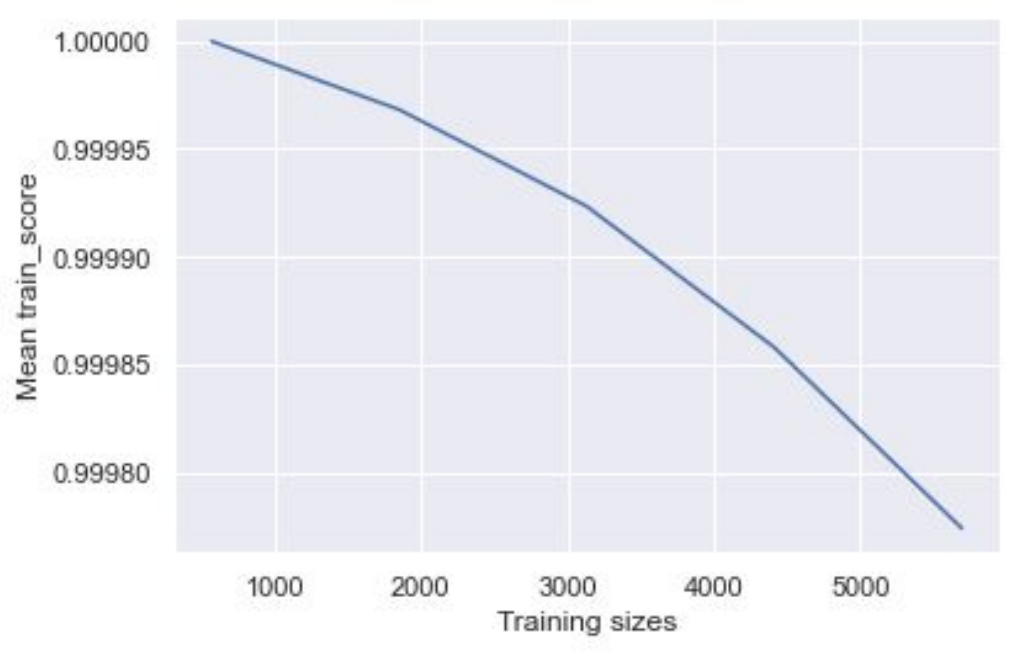


Accuracy:

0.575084467803466



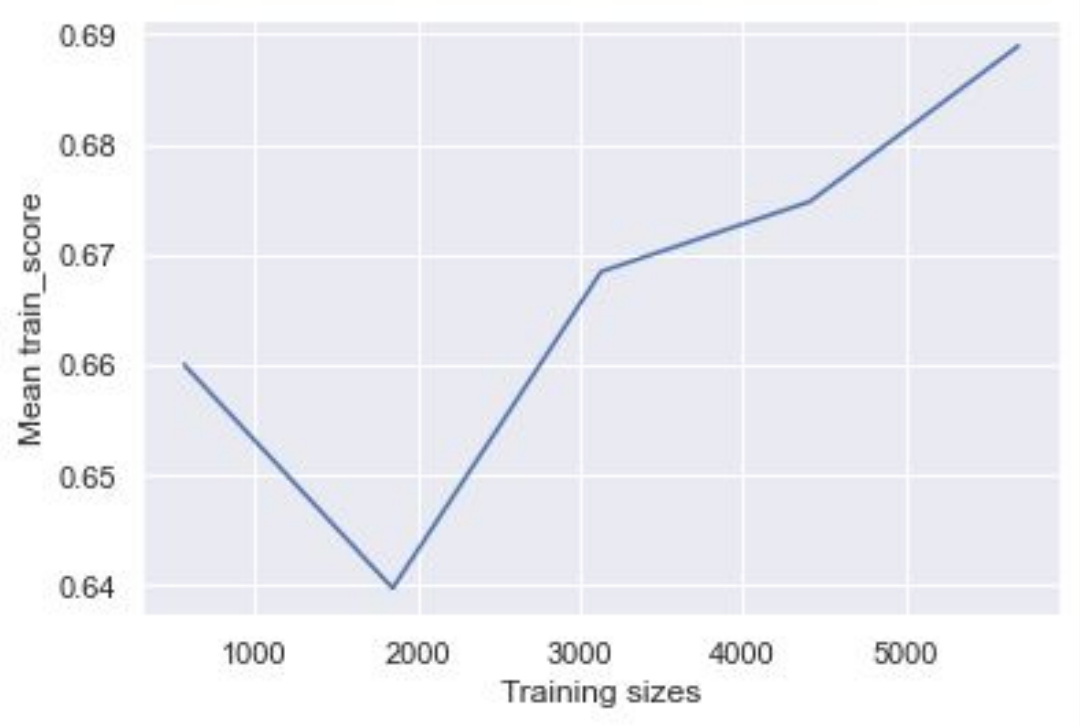
Decision Tree Training Model



Accuracy:

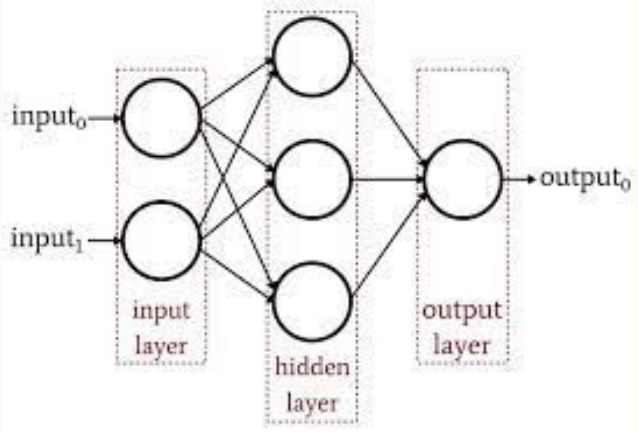
0.7294516932983184

Multilayer Perceptron (MLP) Training Model

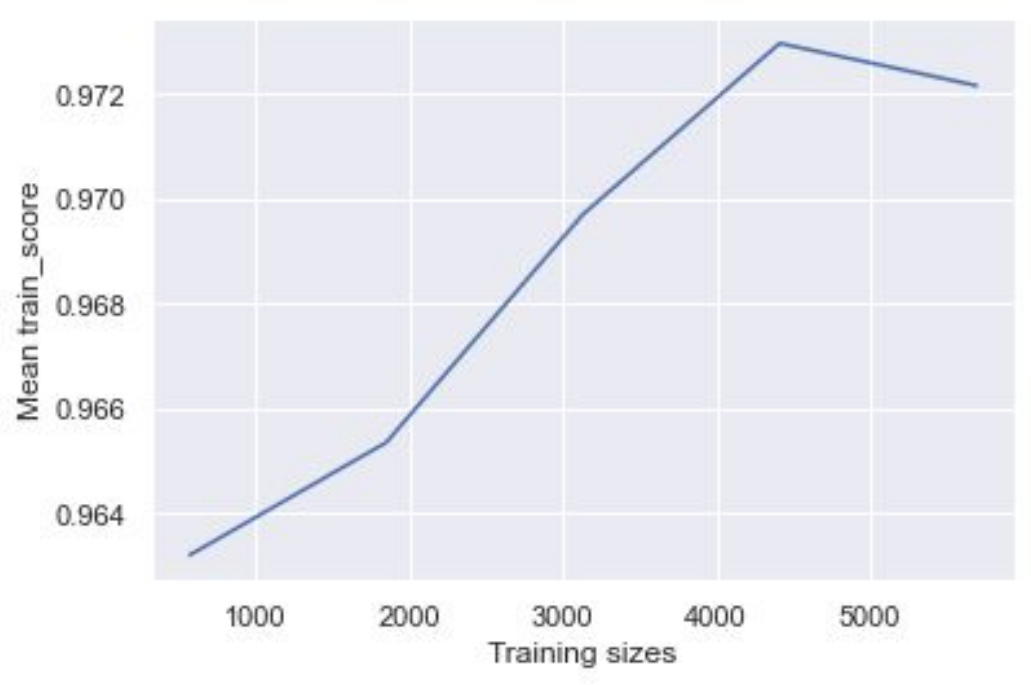


Accuracy:

0.6930506508632484



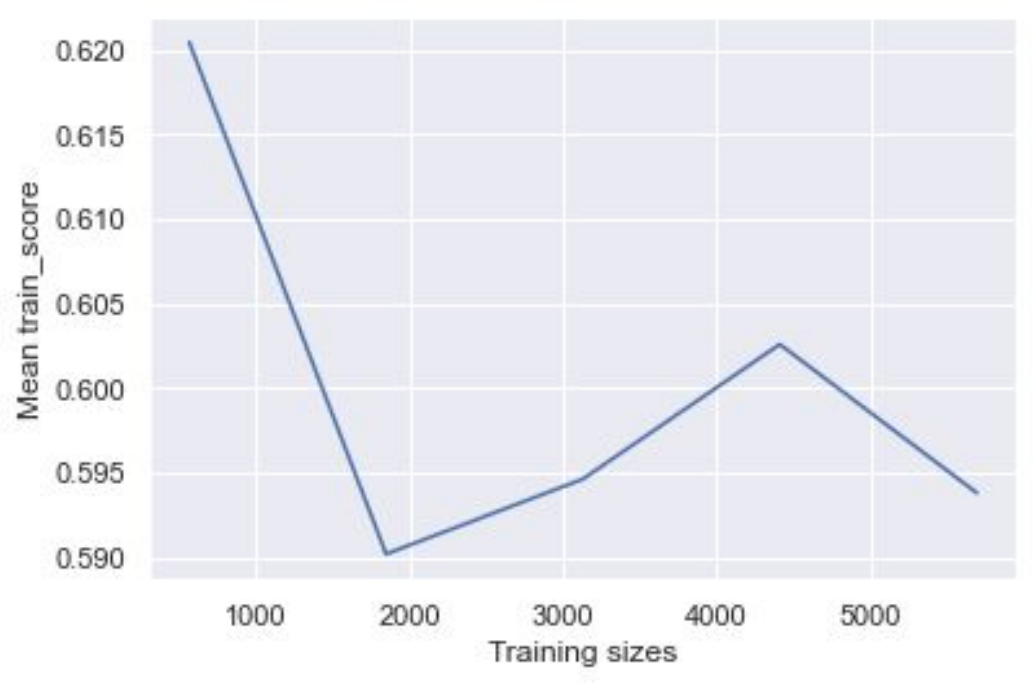
Random Forest Training Model



Accuracy:

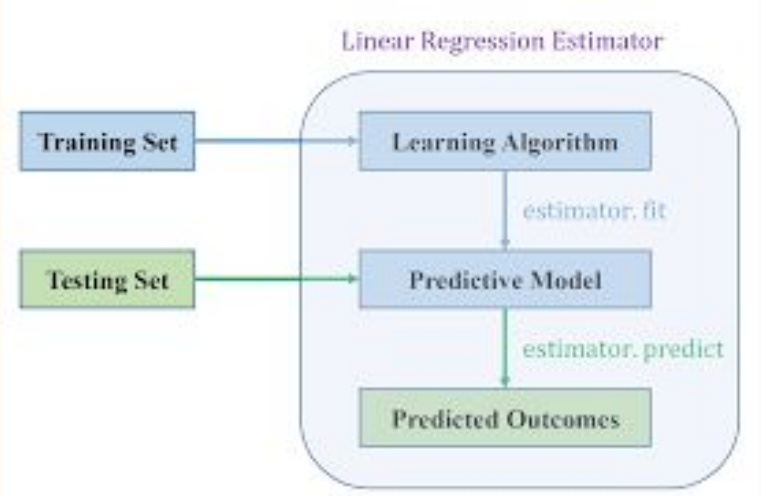
0.8502954380903031

Linear Regression Training Model

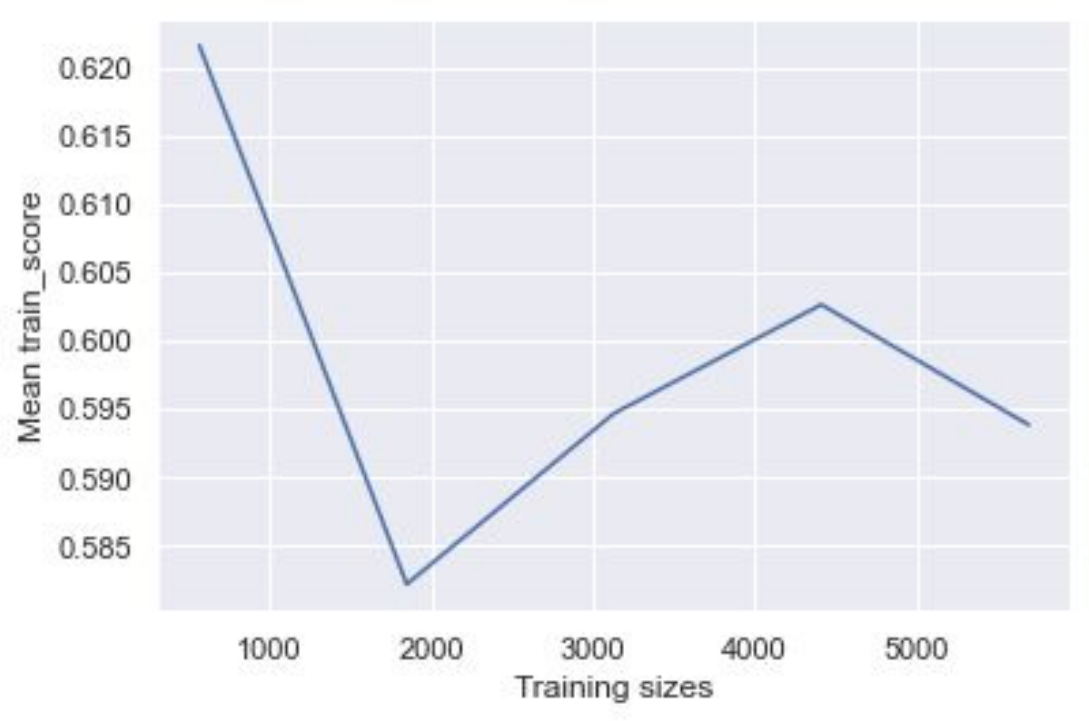


Accuracy:

0.5890179004546502



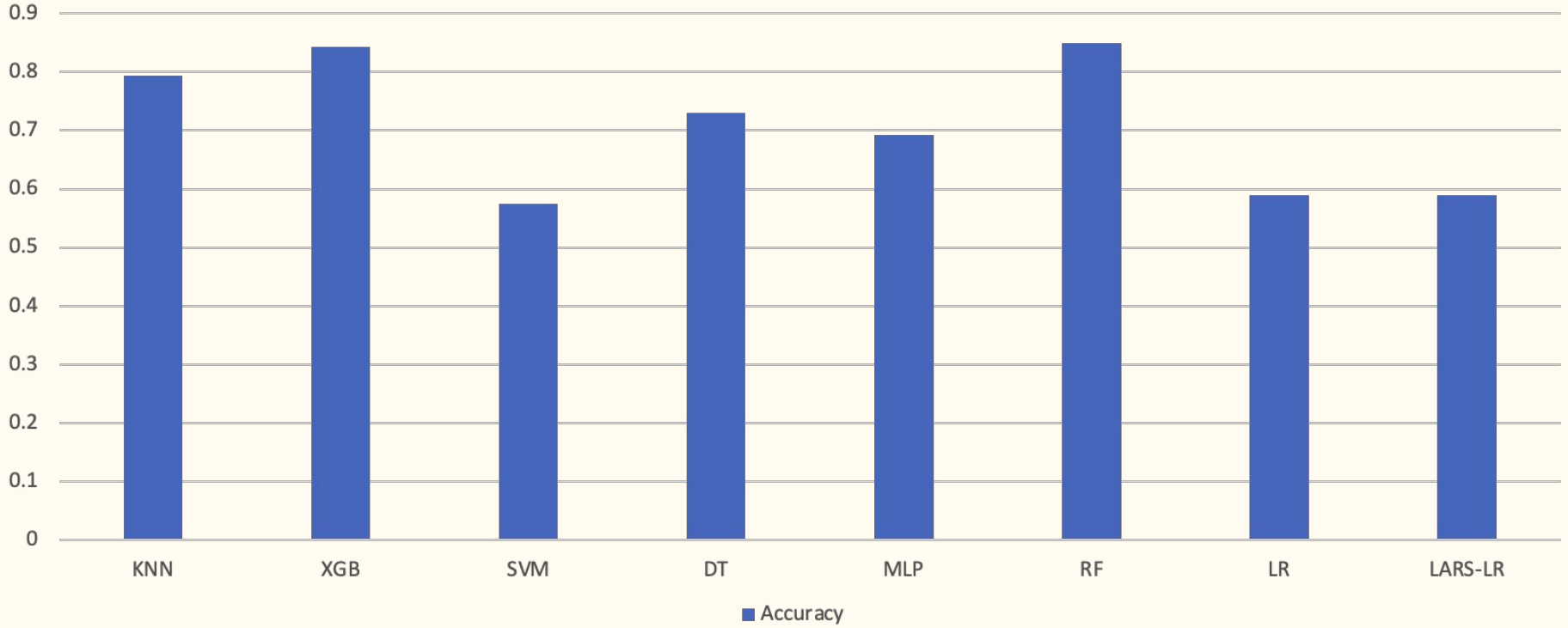
LARS Linear Regression Model



Accuracy:

0.5890383006435393

Summary


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
KNN
XGB
SVM
DT
MLP
RF
LR
LARS-LR
Accuracy

Thank you! Questions?