

## ALGORITHM-BASED ESTIMATE

An Alternate Approach  
of Construction Cost  
Prediction in Early  
Stage of Design

Shih-Chi  
Liu  
Hathaway  
Dinwiddie



## DEFINITION

Project Overview  
Problem Statement  
ML Pipeline



## ANALYSIS

Data Exploration  
Exploratory Visualization  
Algorithms & Techniques  
Benchmark



## METHODOLOGY

Data Pre-processing  
Implementation  
Refinement



**SHIH-CHI LIU**

Pre-Construction Engineer



**HATHAWAY DINWIDDIE**  
General Contractor



### ROUGH ORDER OF MAGNITUDE

- Average of similar projects
- Time-efficient
- Less accurate



### DETAILED ESTIMATE

- Line-by-line details
- Time-consuming
- More accurate



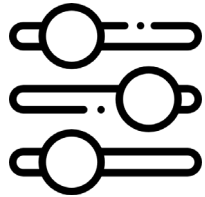
### ROUGH ORDER MAGNITUDE

- Average of similar projects
- **Time-efficient**
- Less accurate



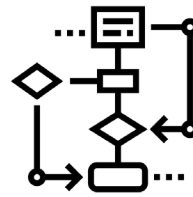
### DETAILED ESTIMATE

- Line-by-line details
- Time-consuming
- **More accurate**



**PROJECT PARAMETER**

Numerical  
Non-numerical



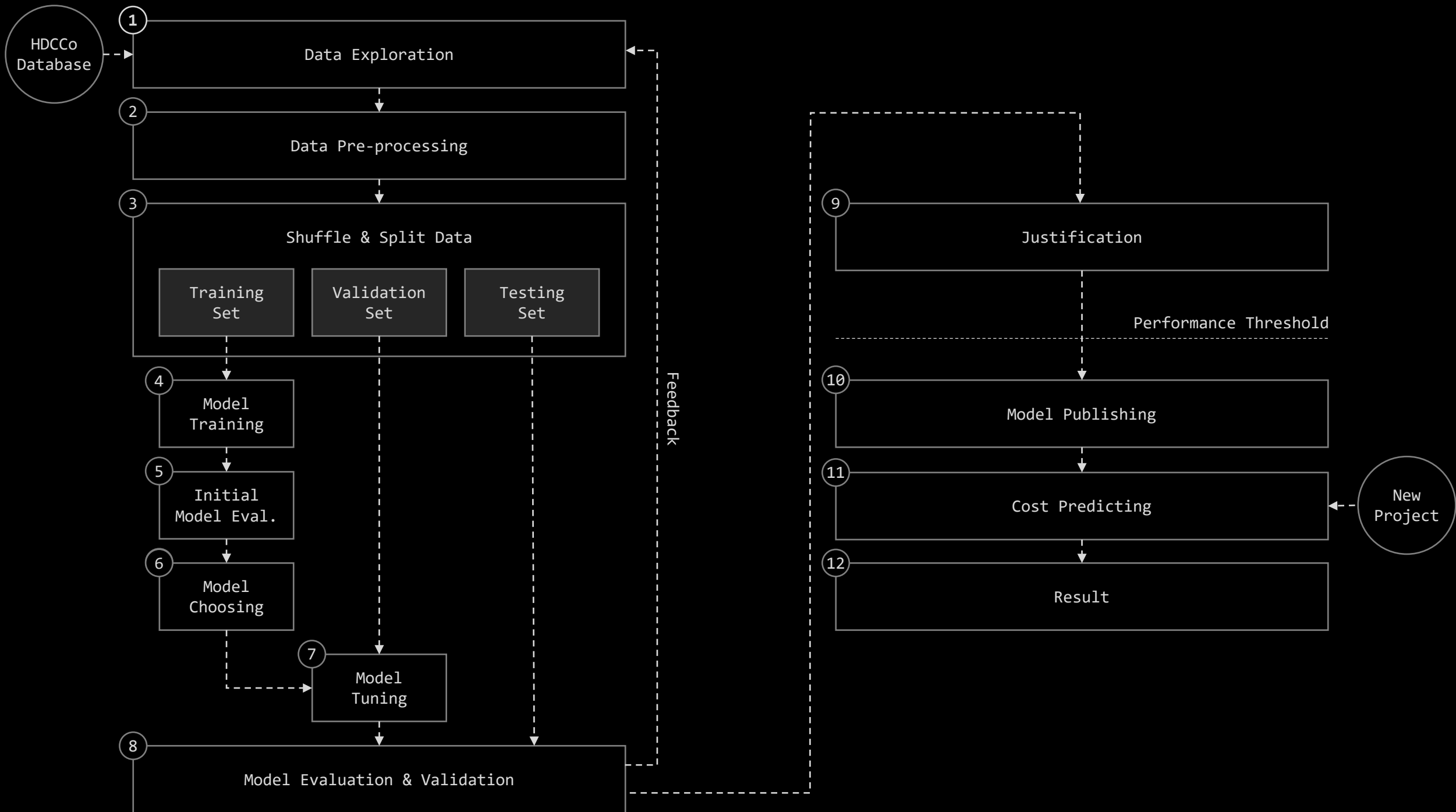
**MACHINE LEARNING**

Regression Model



**CONSTRUCTION COST**

Numerical







## DEFINITION

Project Overview  
Problem Statement  
ML Pipeline



## ANALYSIS

Data Exploration  
Exploratory Visualization  
Algorithms & Techniques  
Benchmark



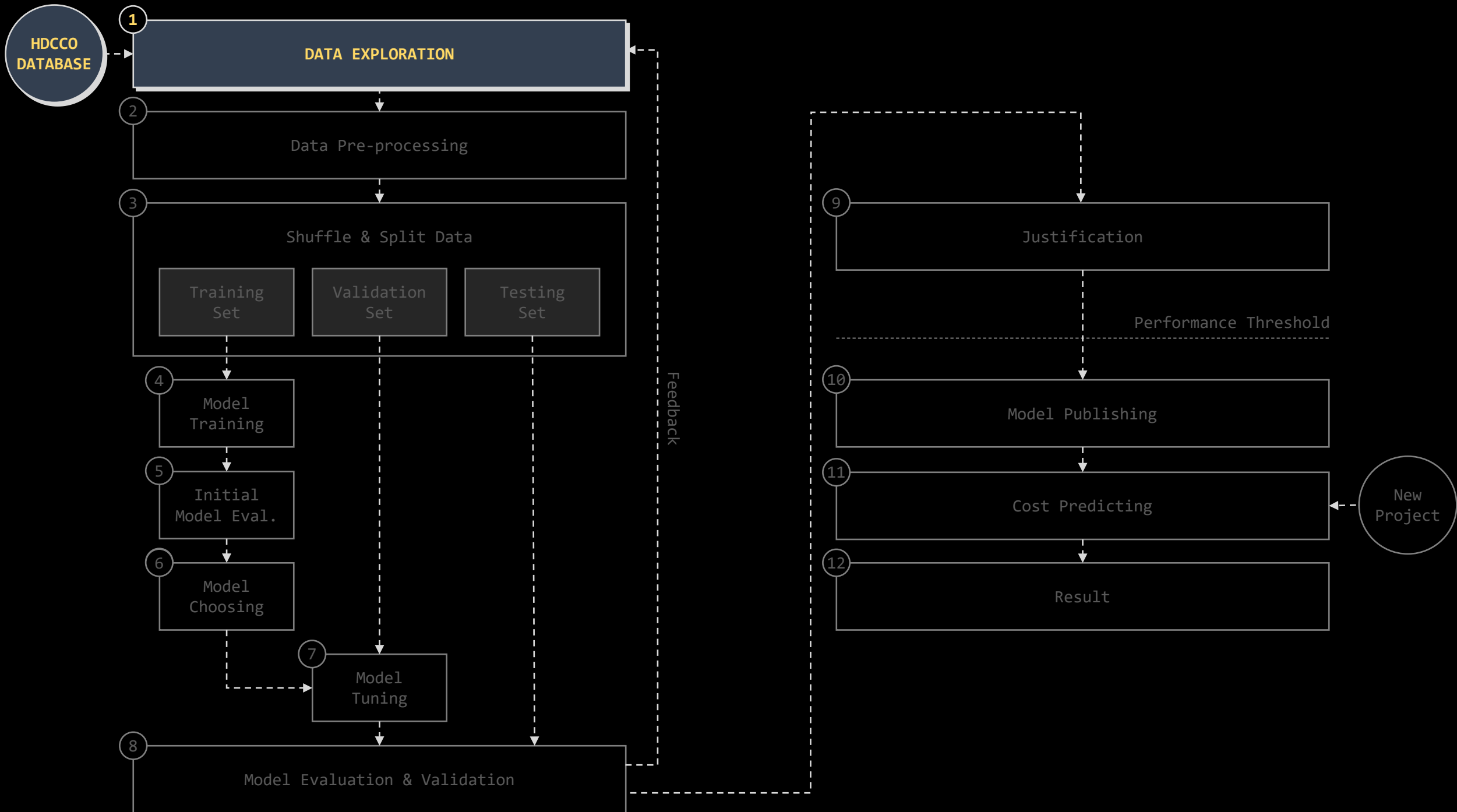
## METHODOLOGY

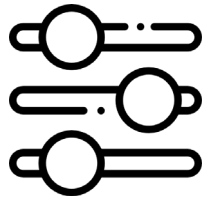
Data Pre-processing  
Implementation  
Refinement



## RESULT

Model Evaluation & Validation  
Justification





(X)

#### NUMERICAL INPUT

- Usable square footage (USF)
- Rentable square footage (RSF)
- Number of floors
- Construction duration

#### NON-NUMERICAL INPUT

- Client type
- Project scope
- Project function
- Project form
- Project location
- Project environment
- Logistic condition
- Design method
- Fit & finish
- Architecture system

```
# Display a description of the dataset
display(data.describe())
```

	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
<b>count</b>	1.170000e+02	115.000000	115.000000	66.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
<b>mean</b>	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
<b>std</b>	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
<b>min</b>	4.575870e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
<b>25%</b>	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
<b>50%</b>	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
<b>75%</b>	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
<b>max</b>	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

8 rows × 69 columns

```
# Display a description of the dataset
display(describe())
```

	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
count	1.170000e+02	115.000000	115.000000	66.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
mean	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
std	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
min	4.575870e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
50%	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
75%	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
max	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

8 rows × 69 columns

```
# Display a description of the dataset  
display(data.describe())
```

NUMERICAL INPUT

	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
count	1.170000e+02	115.000000	115.000000	66.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
mean	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
std	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
min	4.575870e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
50%	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
75%	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
max	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

8 rows × 69 columns

```
# Display a description of the dataset  
display(data.describe())
```

NON-NUMERICAL INPUT

	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
count	1.170000e+02	115.000000	115.000000	66.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
mean	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
std	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
min	4.575870e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
50%	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
75%	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
max	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

8 rows × 69 columns

```
# Display a description of the dataset
display(data.describe())
```

	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
count	1.170000e+02	115.000000	115.000000	66.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
mean	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
std	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
min	4.575870e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
50%	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
75%	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
max	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

8 rows × 69 columns



```
# Display a description of the dataset  
display(data.describe())
```

	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
count	1.170000e+02	115.000000	115.000000	66.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
mean	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
std	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
min	4.575870e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
50%	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
75%	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
max	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

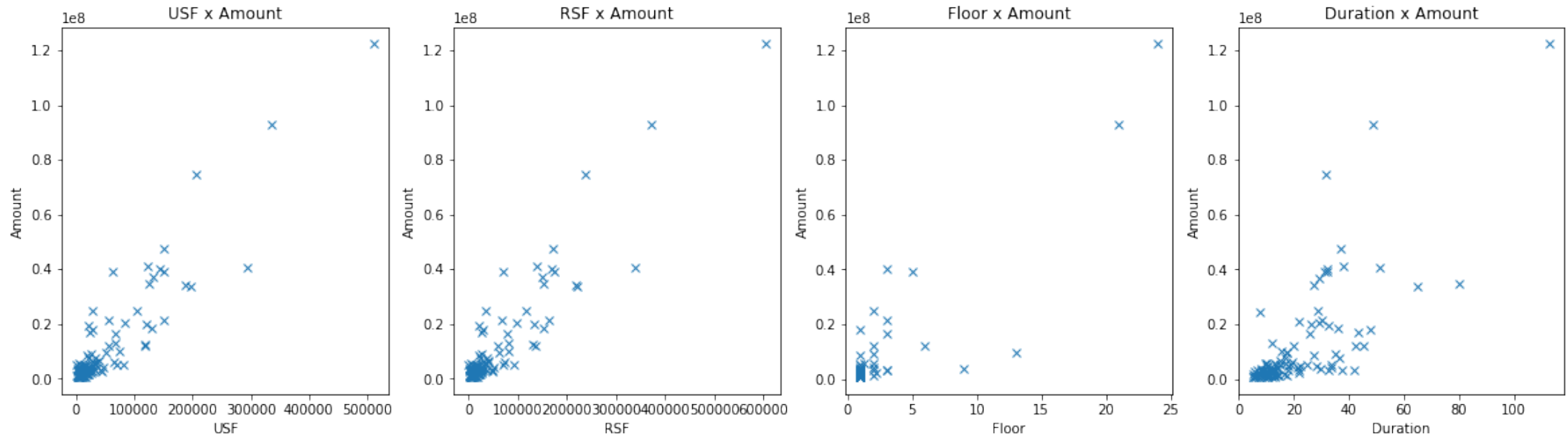
8 rows × 69 columns

```
# Display a description of the dataset
display(data.describe())
```

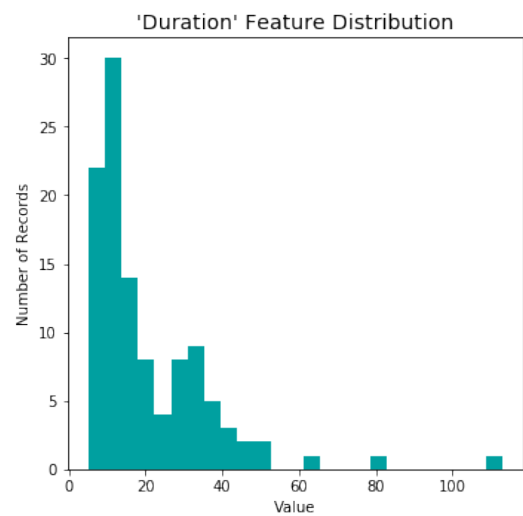
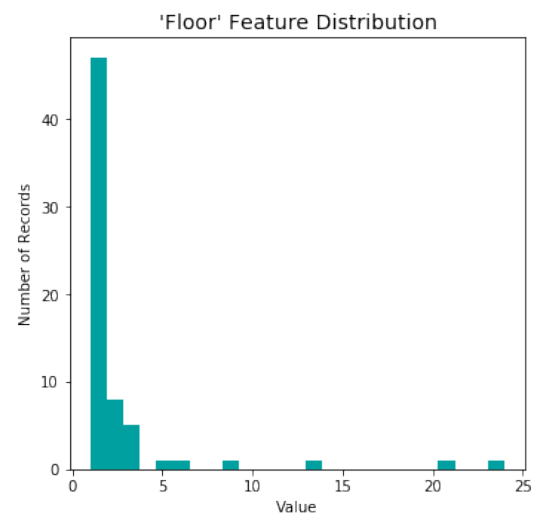
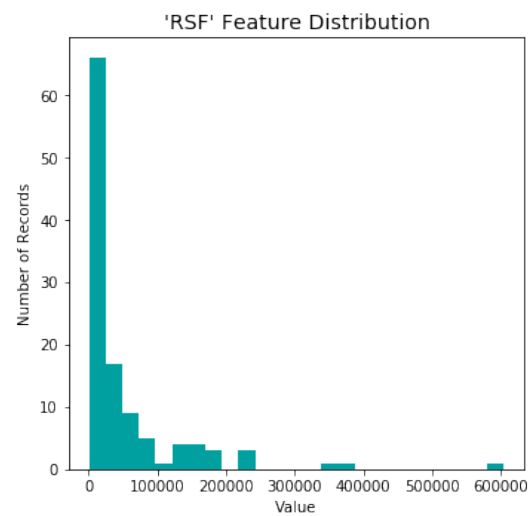
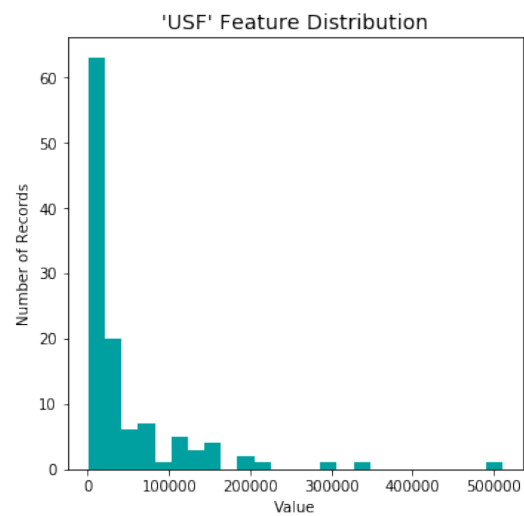
	Amount	USF	RSF	Floor	Duration	2020- Commercial- and- Professional- Services	2530- Consumer- Services	2530.10.20- Hotels- and- Resorts	2530.10.40- Restaurants	2540- Media	...	01- Traditional- Overhead- MEP- Distribution
count	1.170000e+02	115.000000	115.000000	95.000000	110.000000	117.000000	117.000000	117.000000	117.000000	117.000000	...	117.000000
mean	1.056734e+07	46599.556522	53348.704348	2.371212	20.620818	0.128205	0.042735	0.076923	0.017094	0.017094	...	0.940171
std	1.804891e+07	73419.627208	84922.651311	4.080970	16.269743	0.335756	0.203129	0.267615	0.130179	0.130179	...	0.238190
min	4.575671e+05	660.000000	759.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	1.796400e+06	8769.500000	10259.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
50%	3.616244e+06	18437.000000	21242.000000	1.000000	14.300000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
75%	9.656404e+06	48134.000000	56207.500000	2.000000	28.900000	0.000000	0.000000	0.000000	0.000000	0.000000	...	1.000000
max	1.223393e+08	511014.000000	604559.000000	24.000000	113.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000

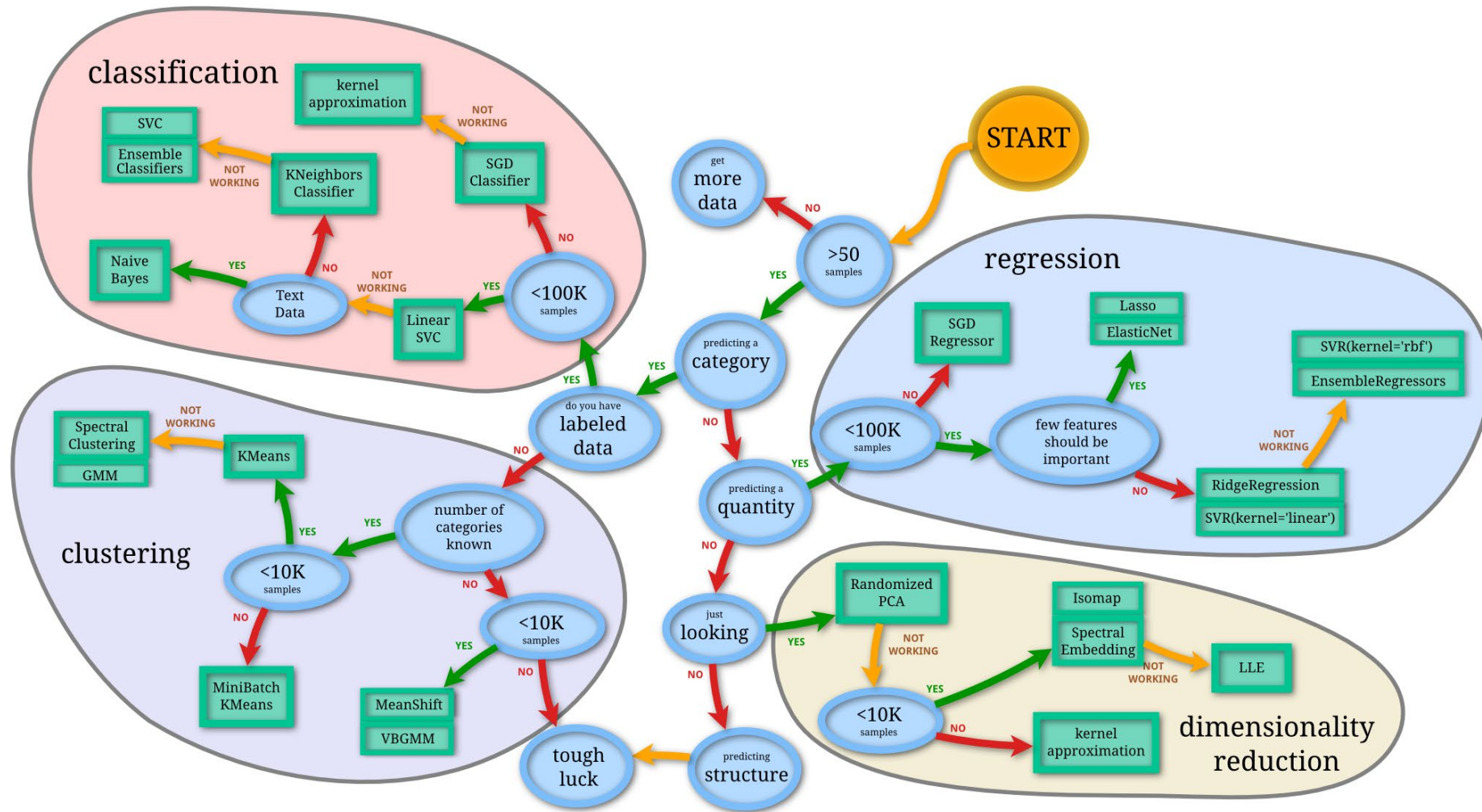
COEFFICIENT OF VARIATION

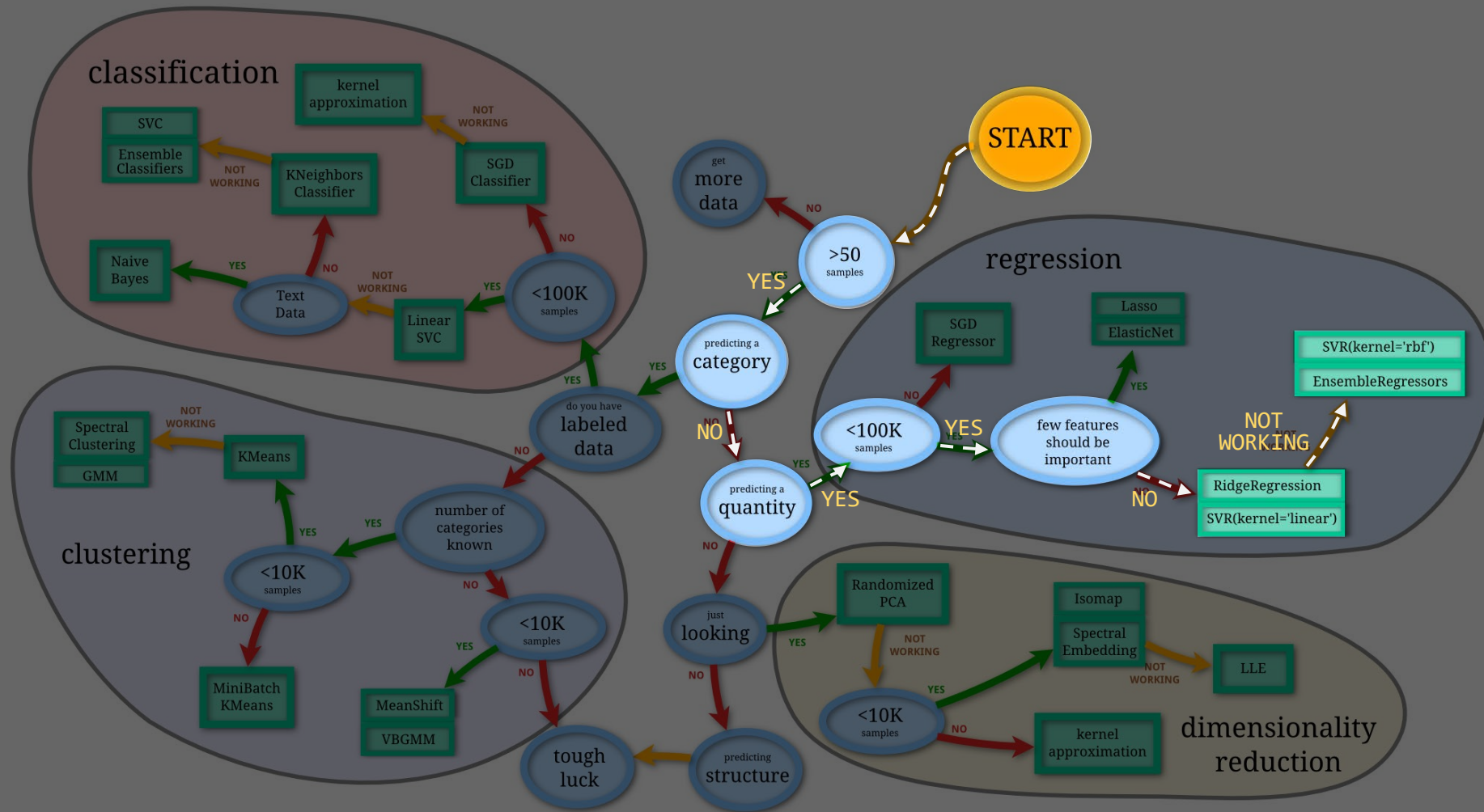
8 rows × 69 columns



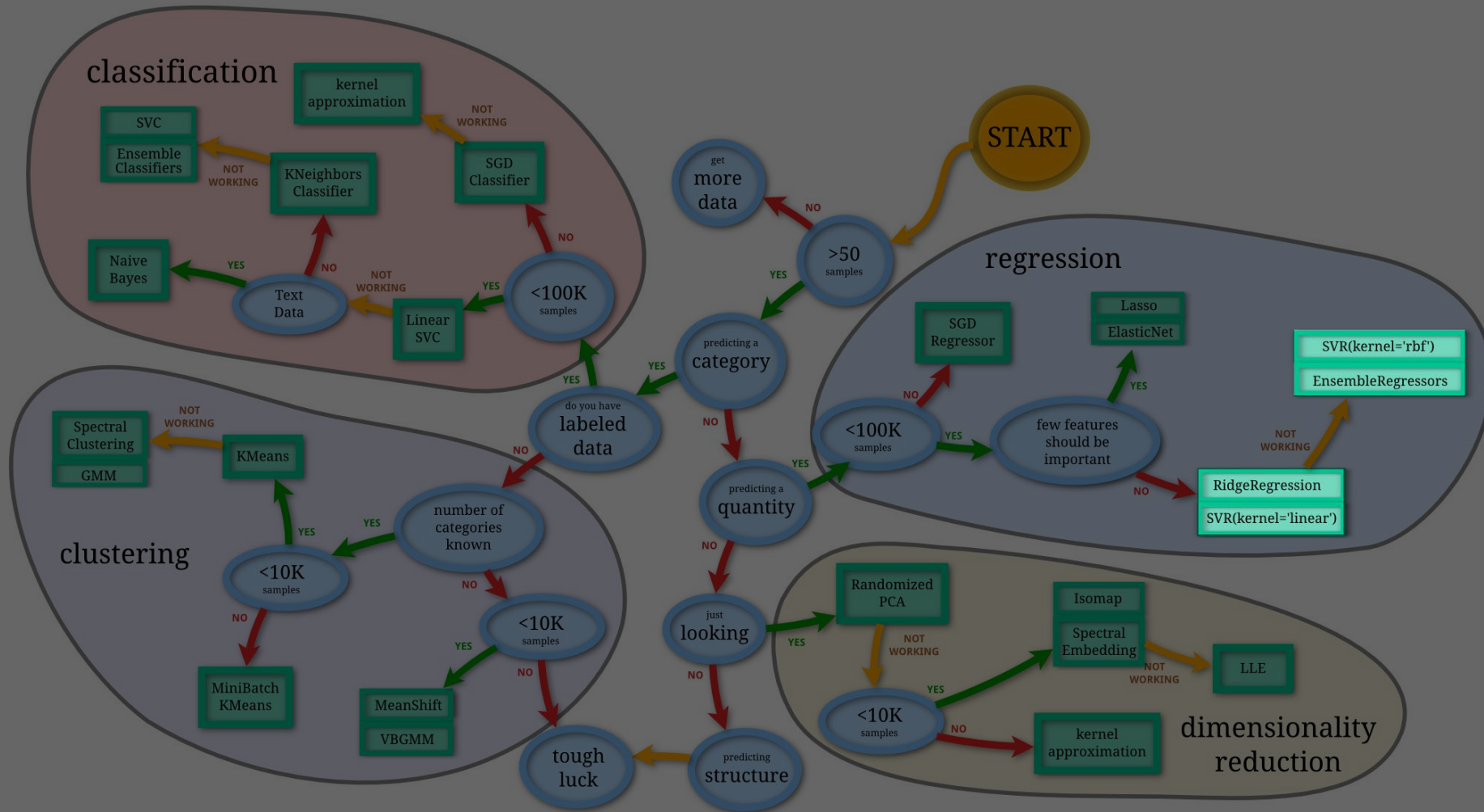
### Skewed Distributions of Continuous Data Features



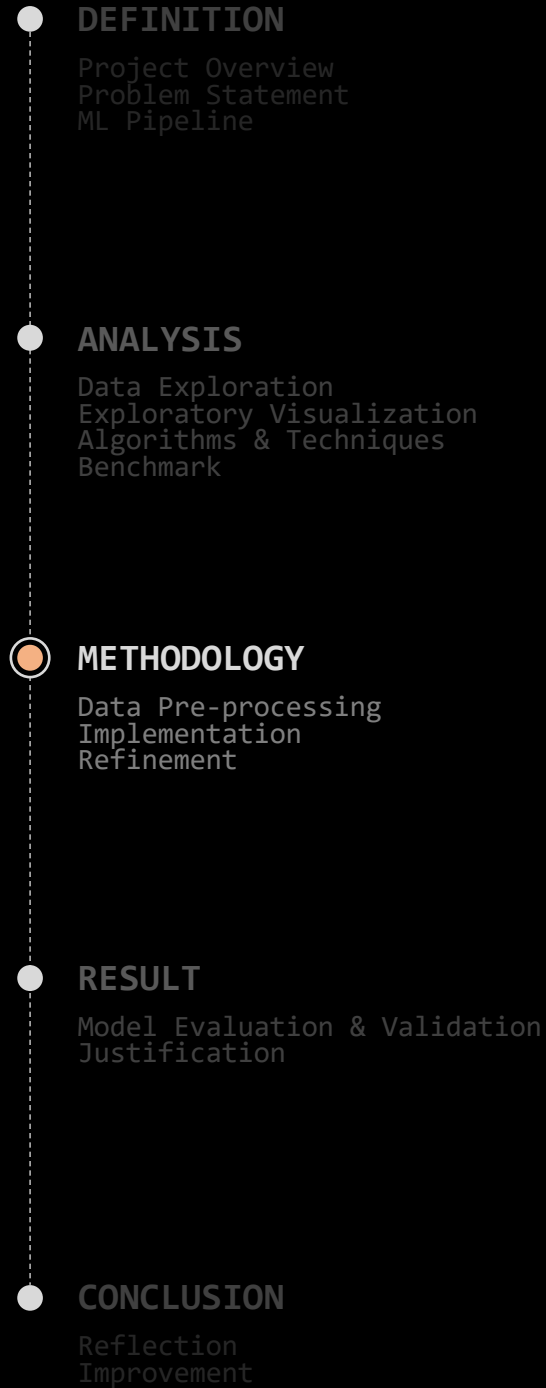




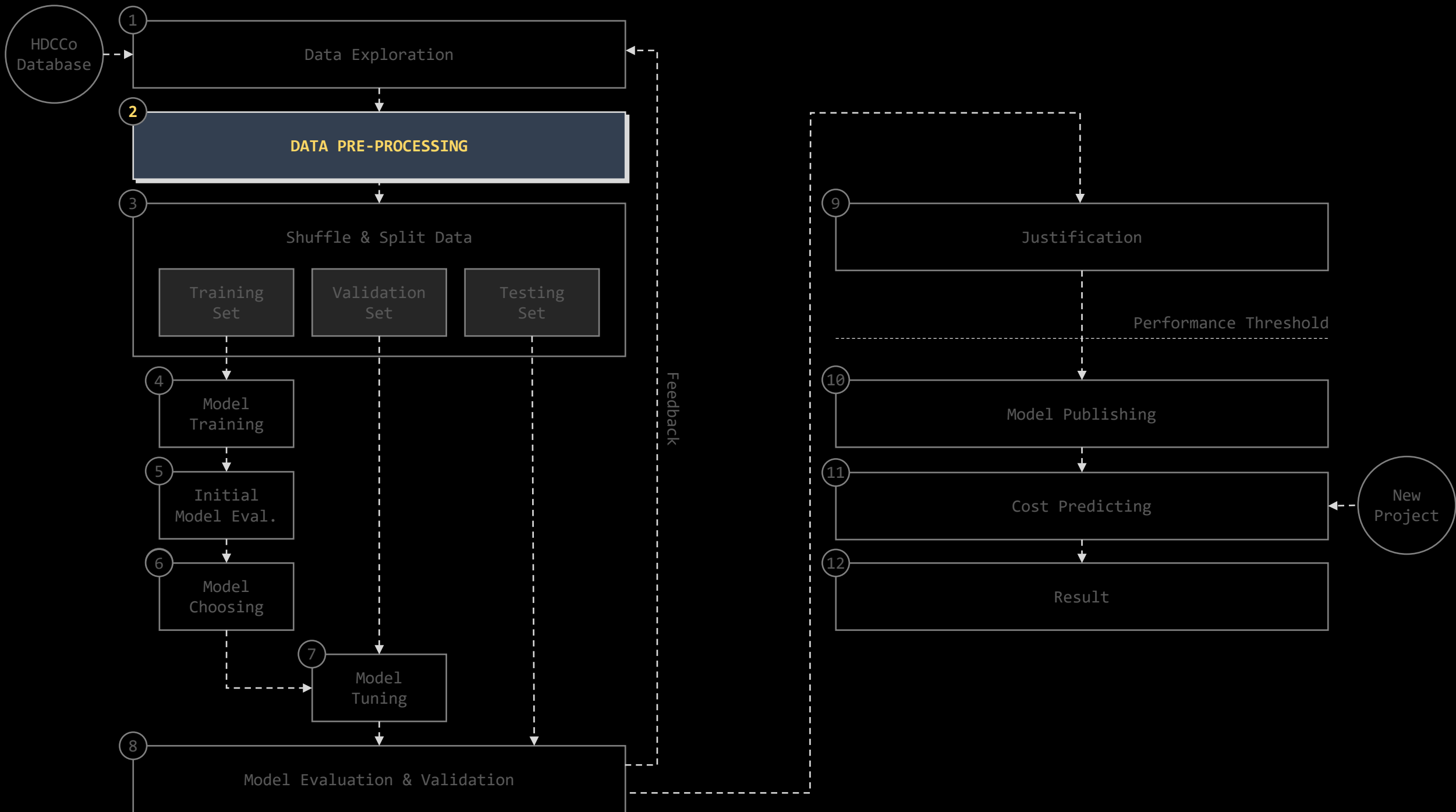




- K Neighbor
- Random Forest
- Xgboost
- LightGBM









Building cost index  
Construction cost index



Bid index



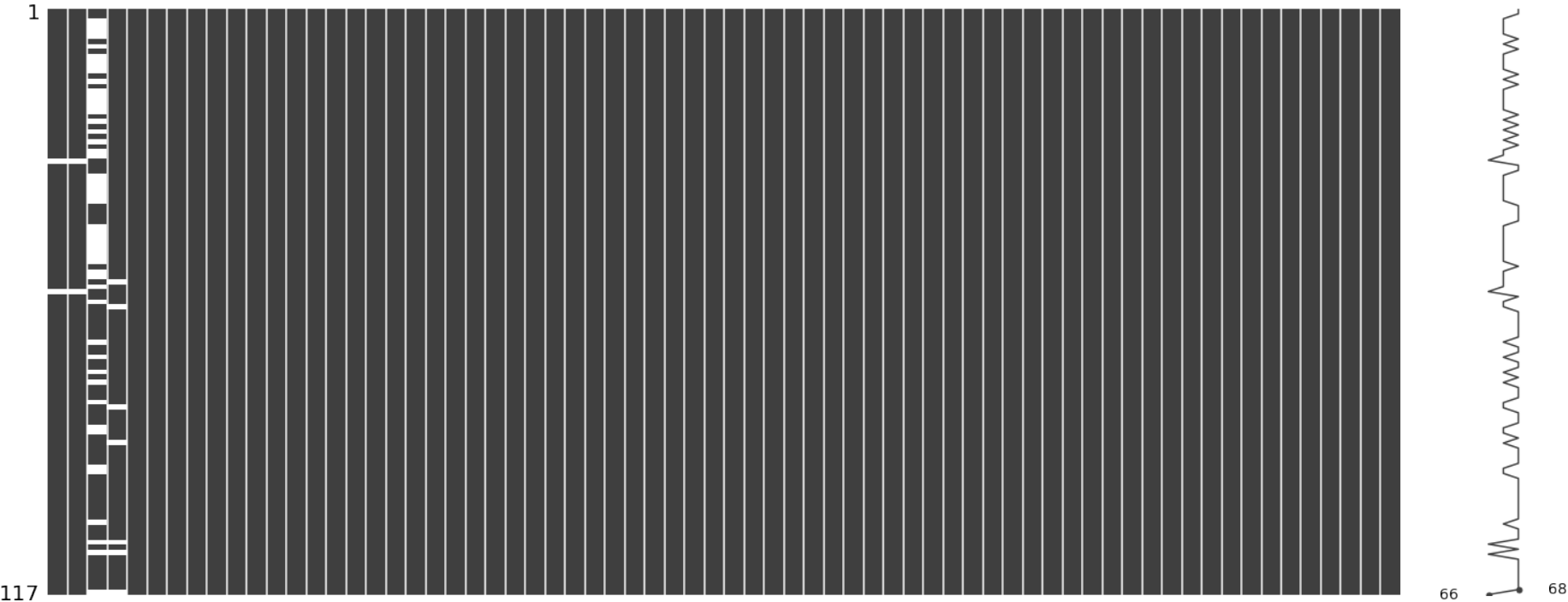
Building cost index



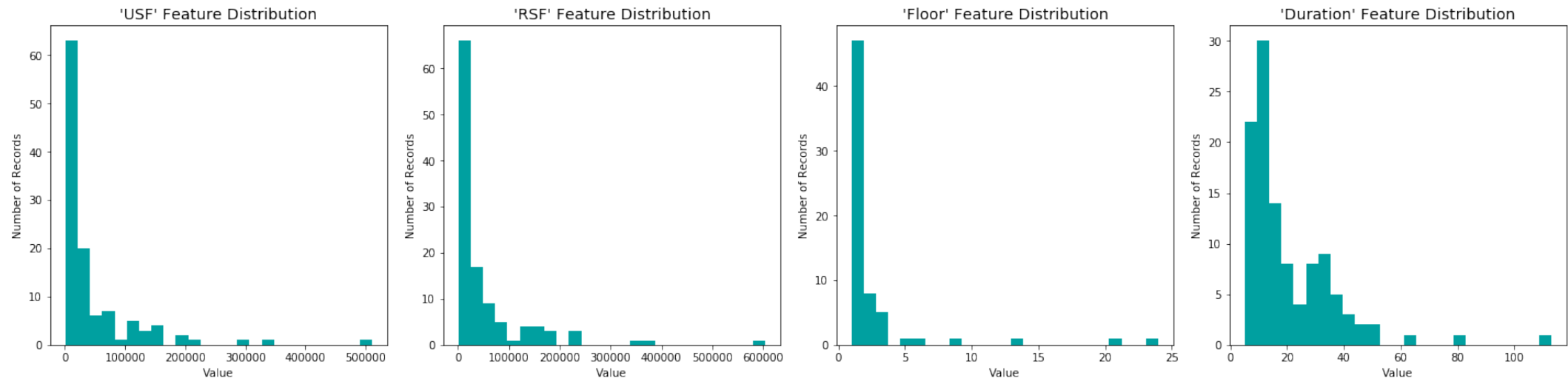
Construction index



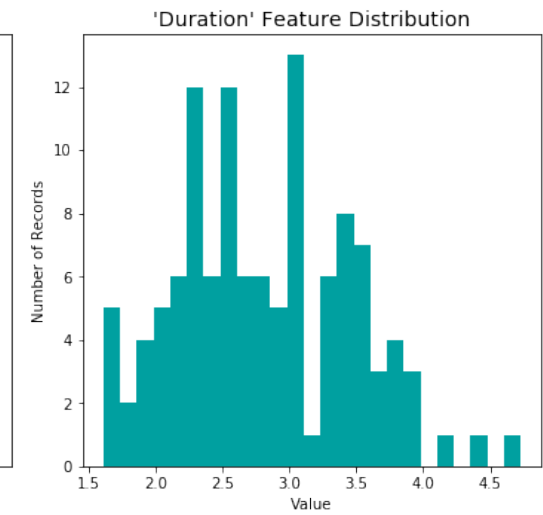
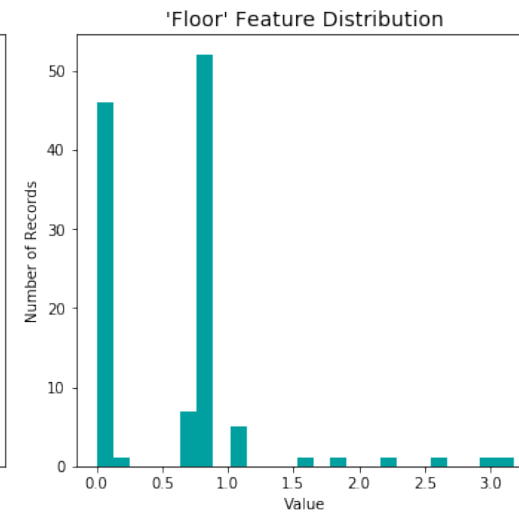
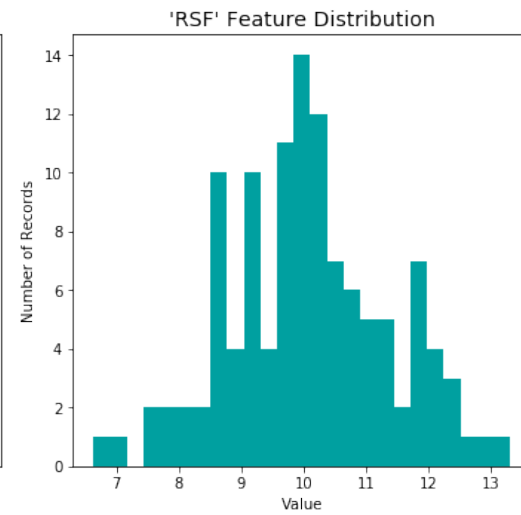
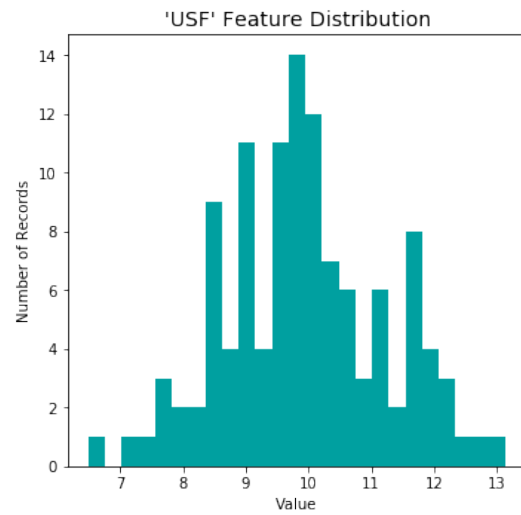
Local subcontractors' labor rates

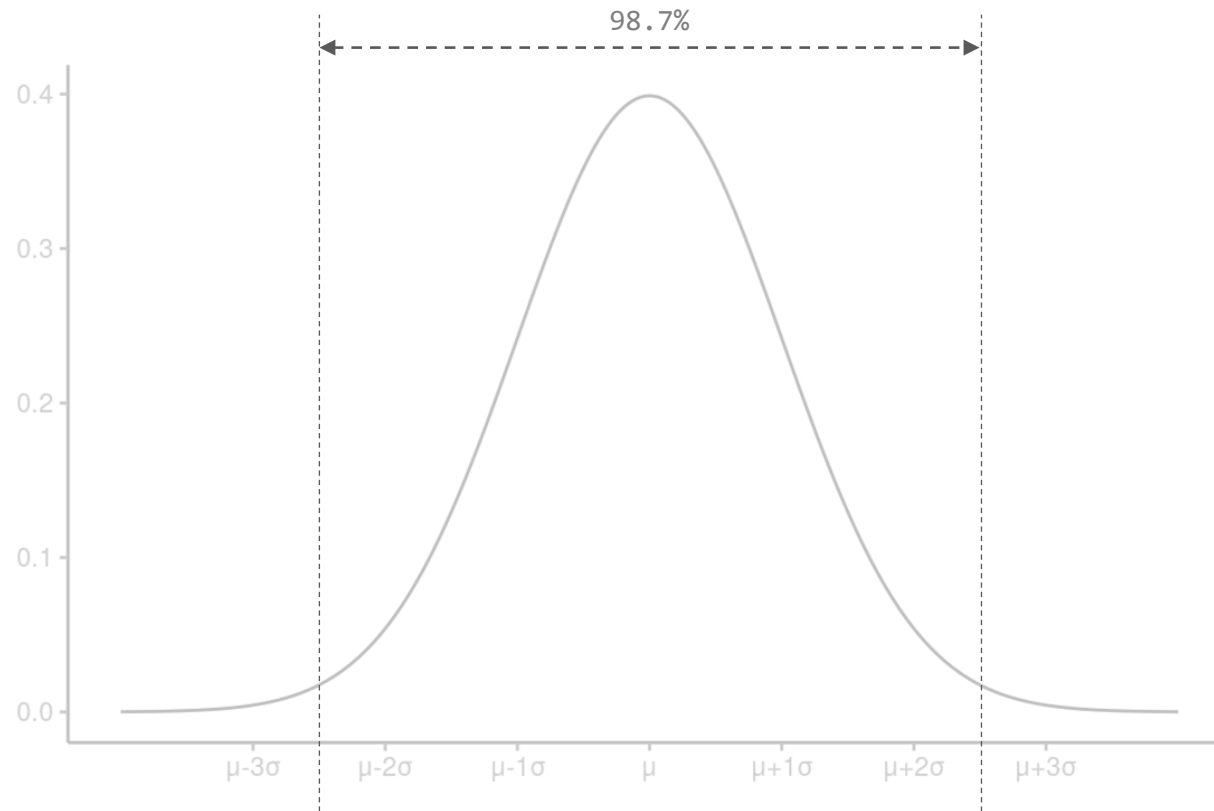


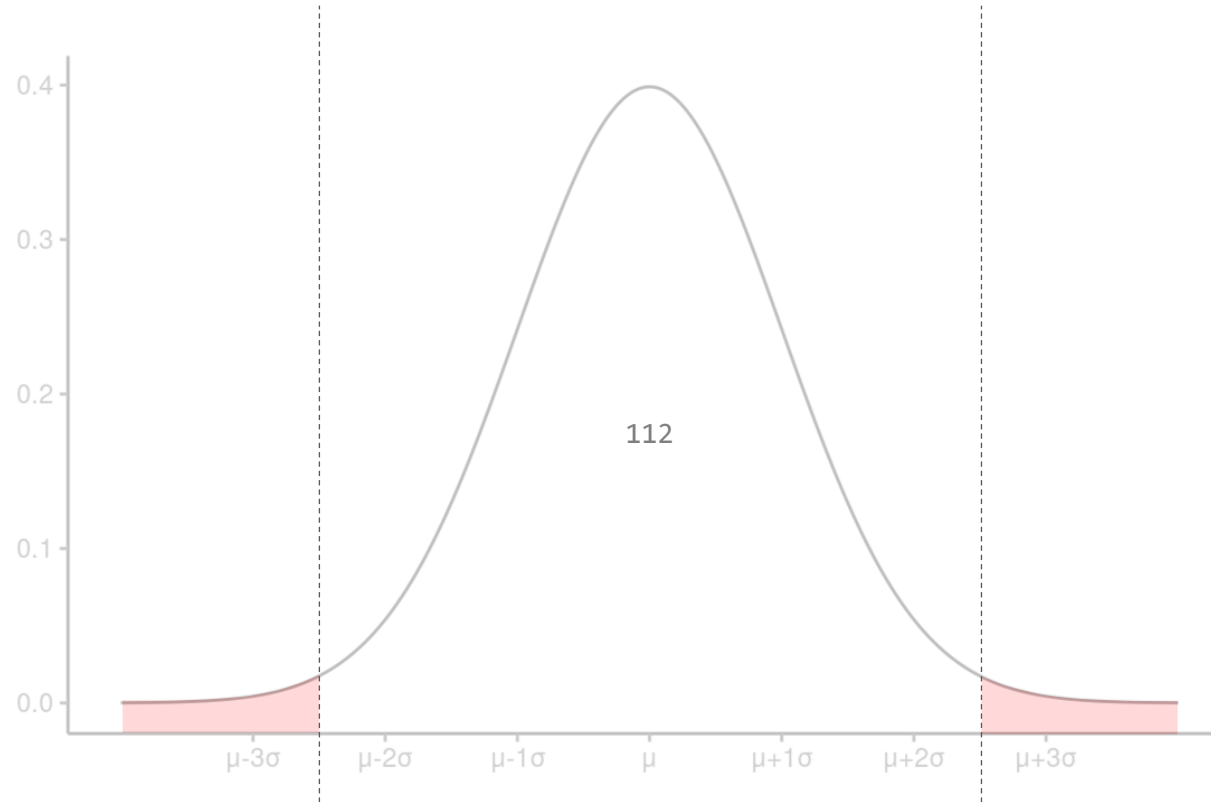
### Skewed Distributions of Continuous Data Features



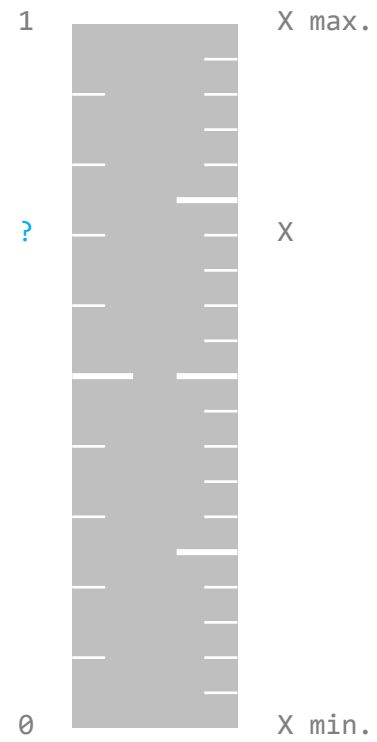
### Log-transformed Distributions of Continuous Data Features



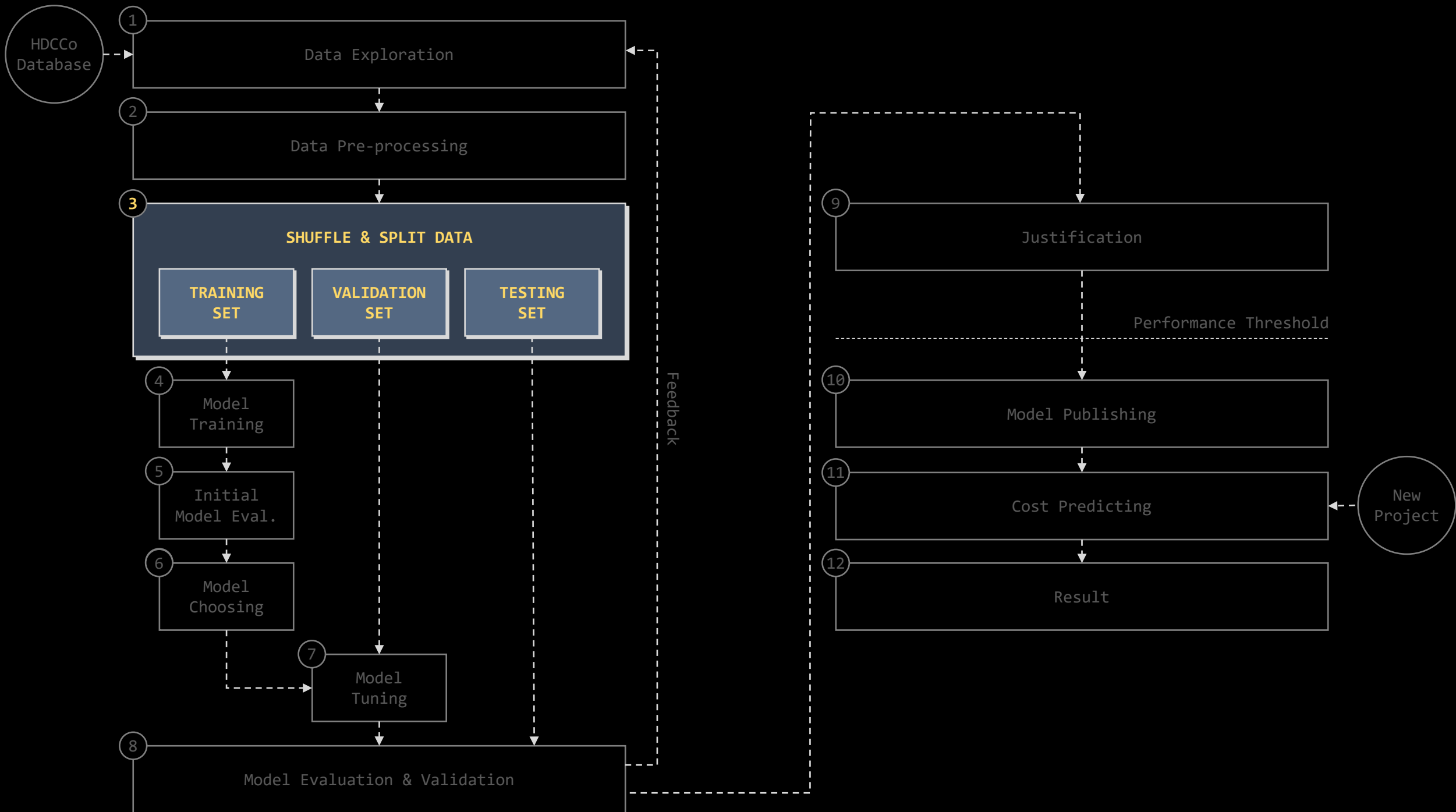




- USF: #42
- RSF: #0 & 42
- Floor: #0, 1, 30, & 56
- Duration: #0







80

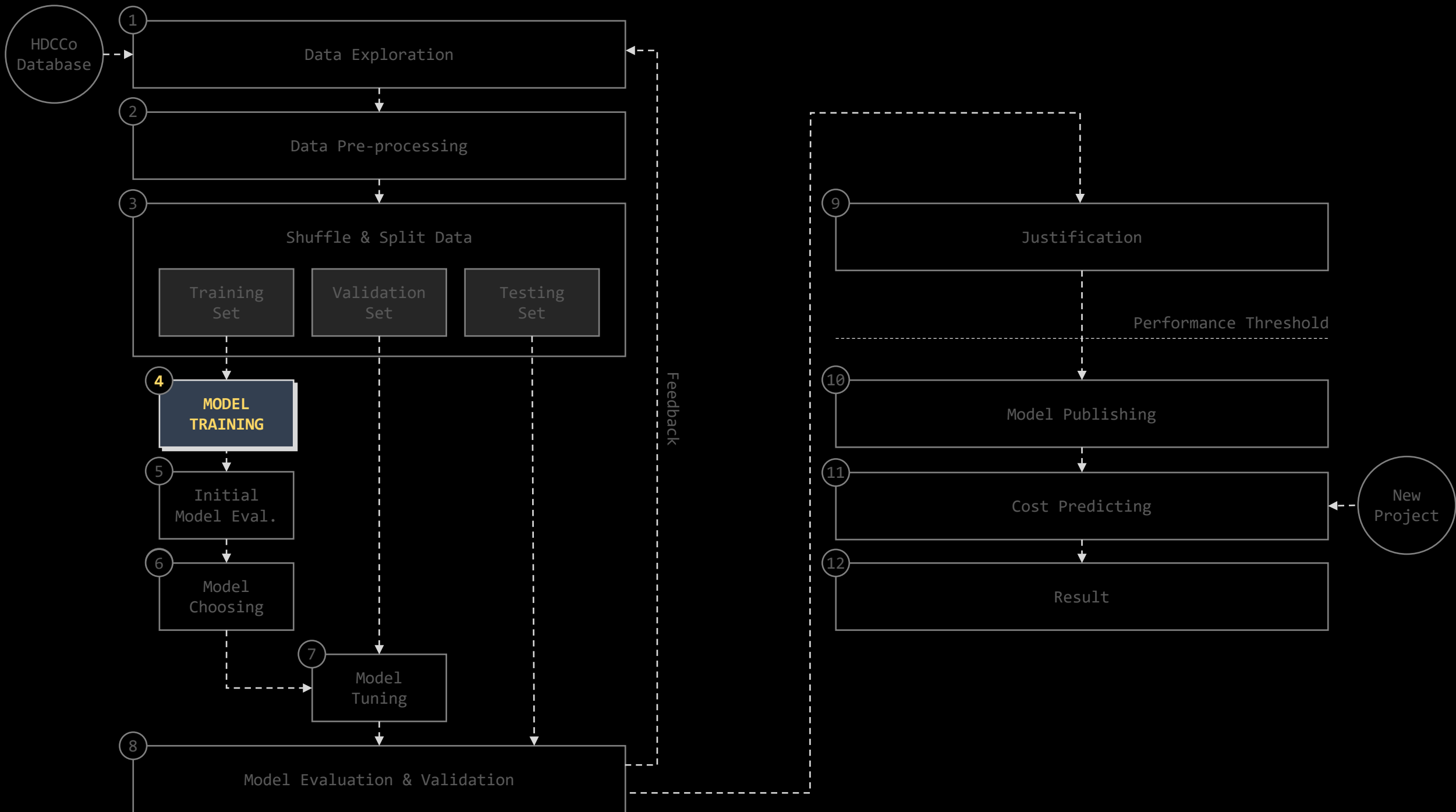
### TRAINING SET

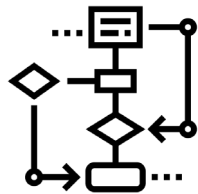
- `X_train`: Feature training set
- `y_train`: Target training set

20

### TESTING SET

- `X_test`: Feature testing set
- `y_test`: Target testing set



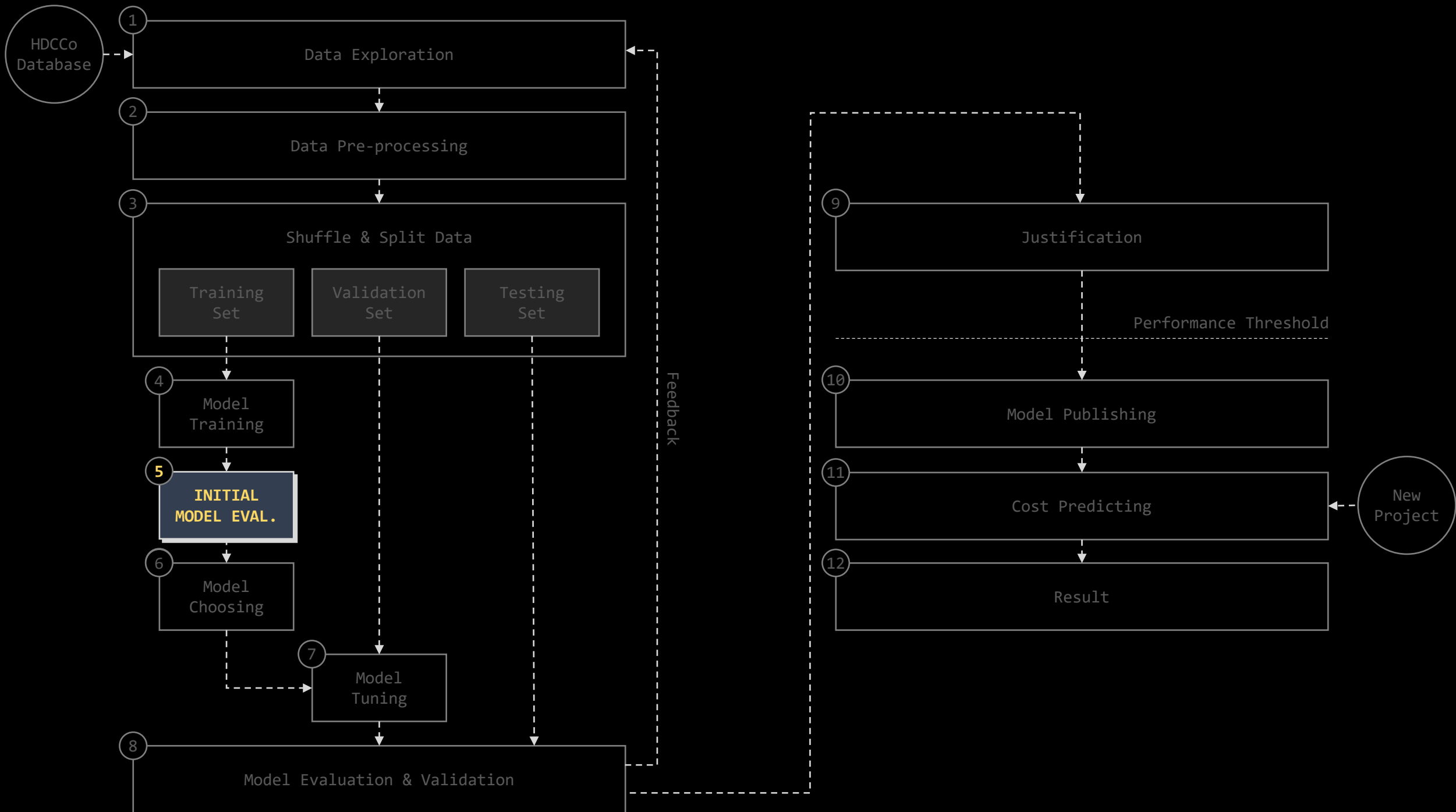


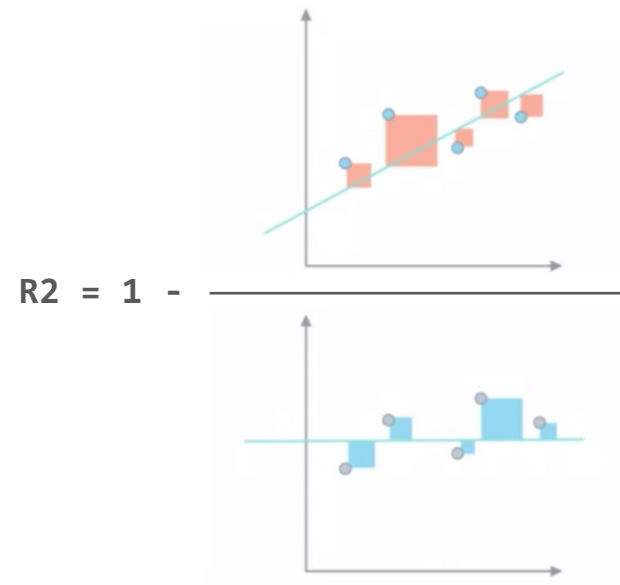
## INPUT

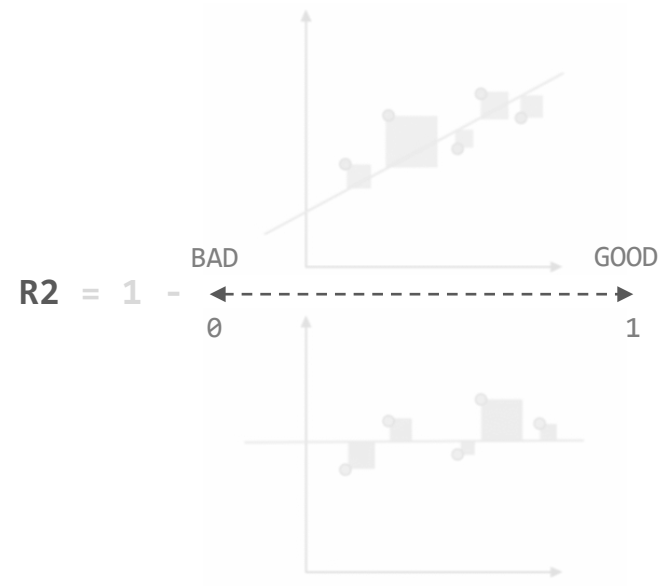
- Learner
  - Ridge
  - RandomForestRegressor
  - SVR
  - KNeighborsRegressor
  - XGBRegressor
  - Lgb
- Sample\_size
- Training\_set
  - X\_train
  - y\_train
- Testing\_set
  - X\_test
  - y\_test

## OUTPUT

- Learner
- training\_size
- R2\_train
- R2\_test

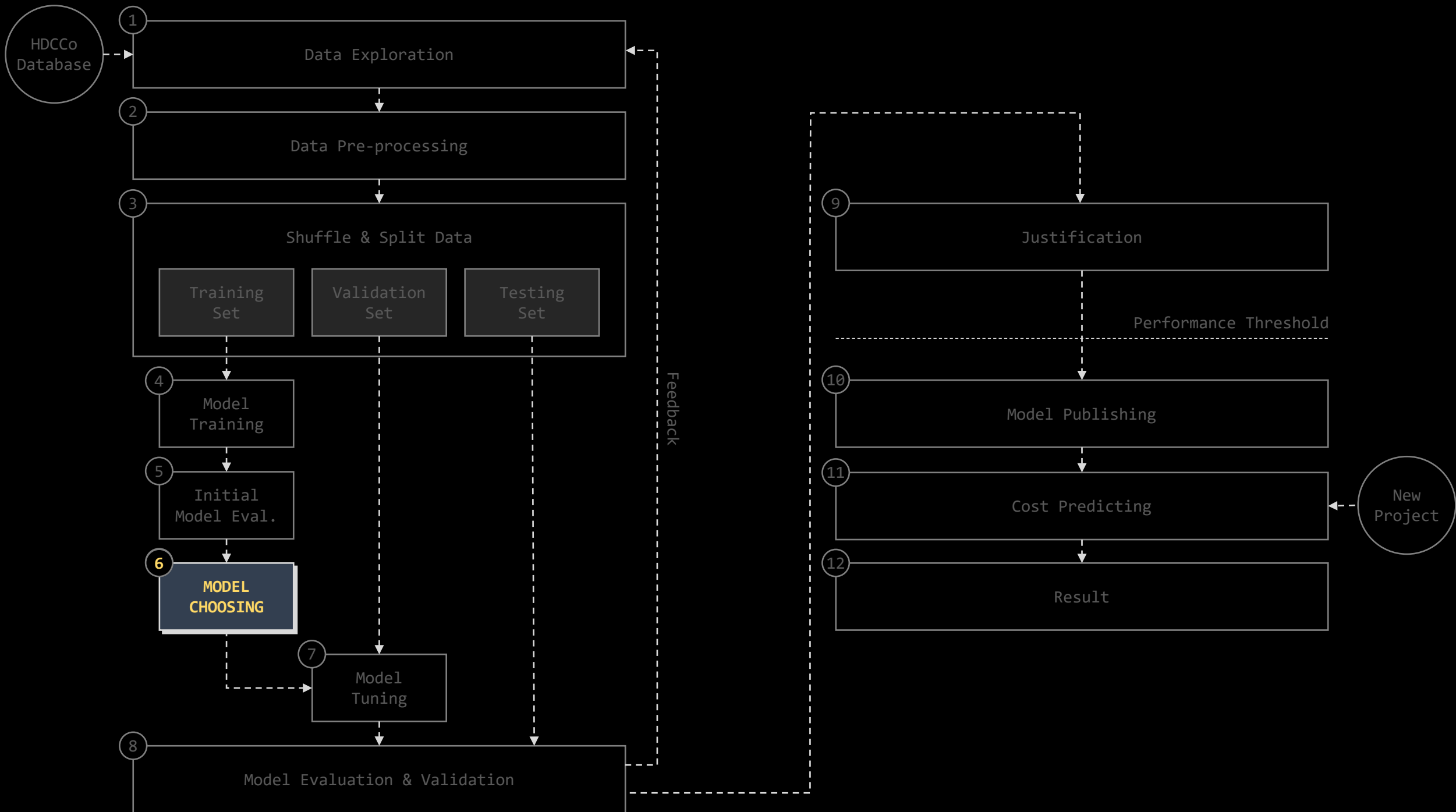




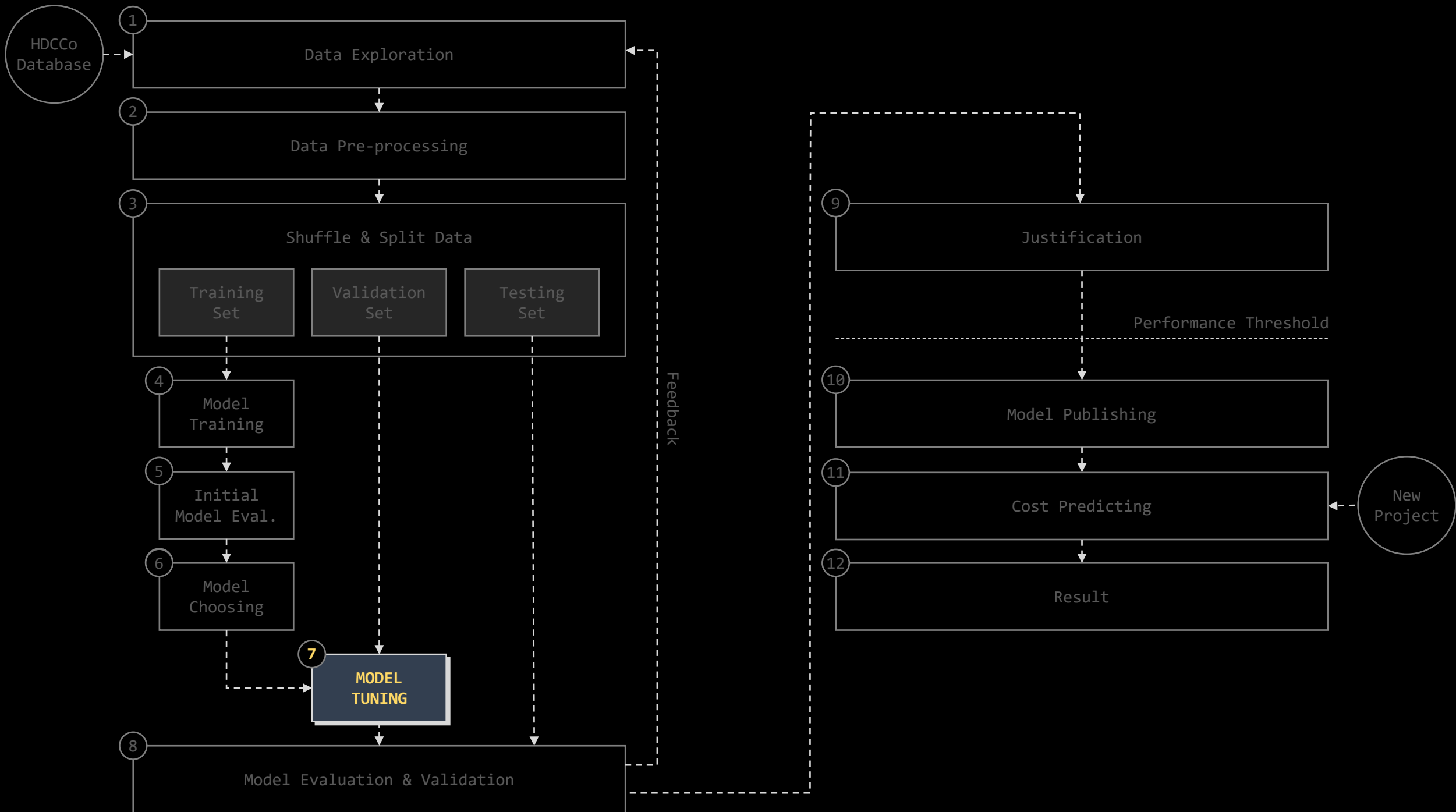


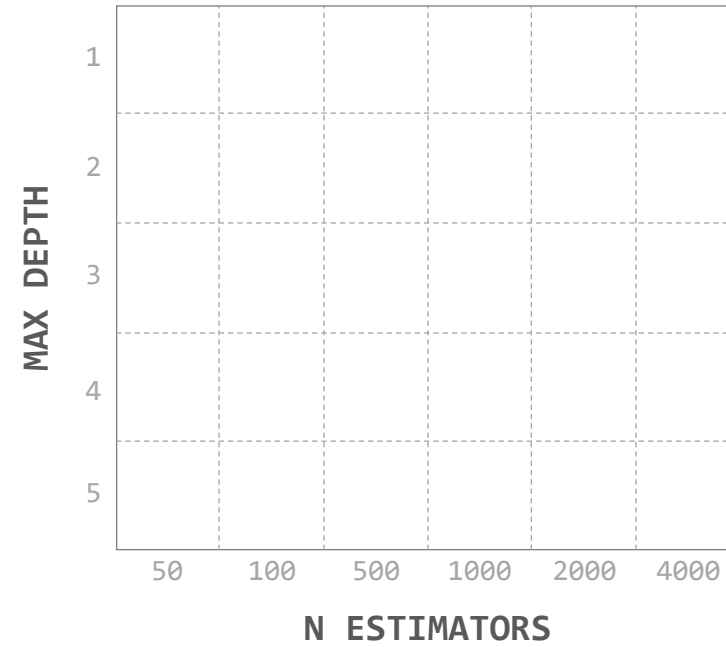
		TRAIN	TEST
R2	• Ridge	0.90	0.66
	• RandomForestRegressor	0.94	0.65
	• SVR, linear	-0.18	-0.19
	• SVR, rbf	-0.18	-0.19
	• KNeighborsRegressor	0.68	0.60
	• XGBRegressor	0.99	0.78
	• Booster	0.84	0.60

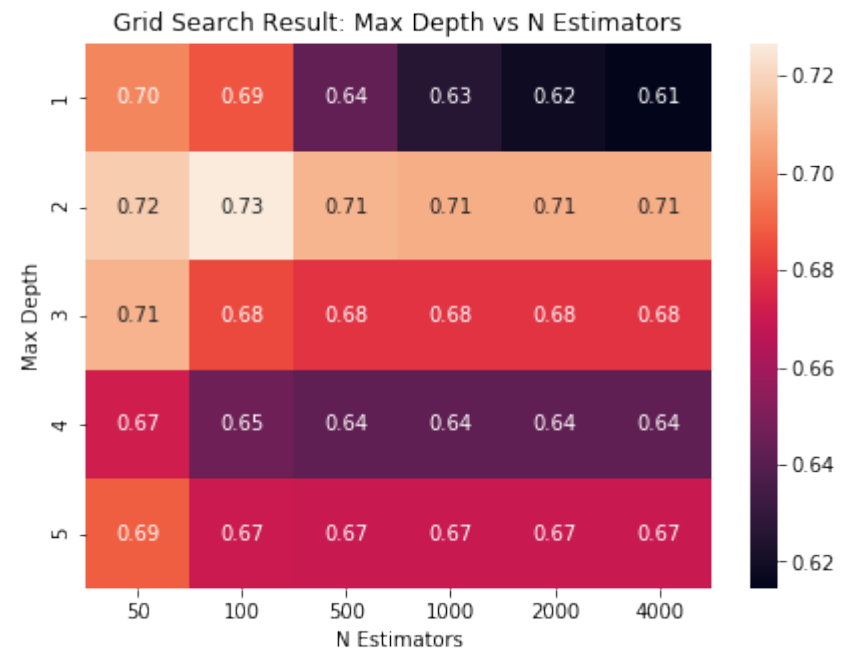


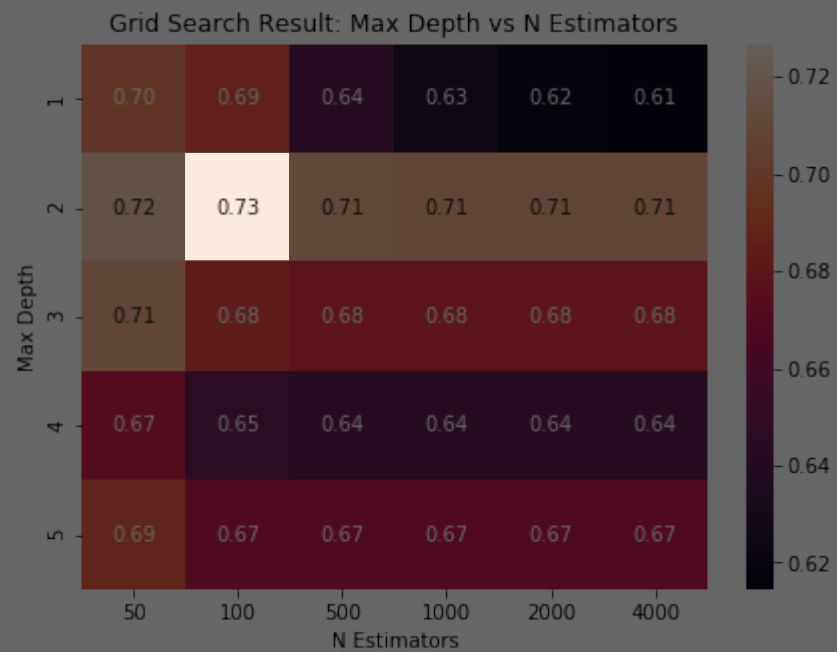


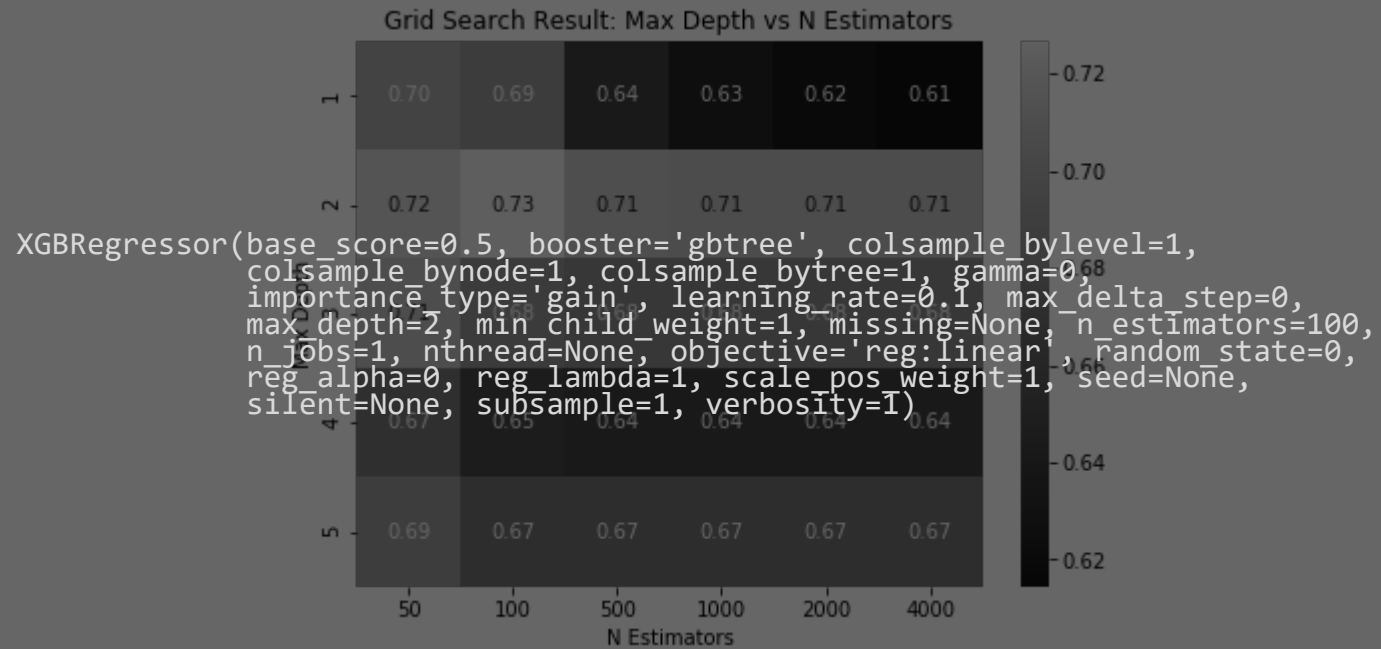
		TRAIN	TEST
R2	• Ridge	0.90	0.66
	• RandomForestRegressor	0.94	0.65
	• SVR, linear	-0.18	-0.19
	• SVR, rbf	-0.18	-0.19
	• KNeighborsRegressor	0.68	0.60
	• <b>XGBRegressor</b>	<b>0.99</b>	<b>0.78</b>
	• Booster	0.84	0.60













## ANALYSIS

Data Exploration  
Exploratory Visualization  
Algorithms & Techniques  
Benchmark

## METHODOLOGY

Data Pre-processing  
Implementation  
Refinement

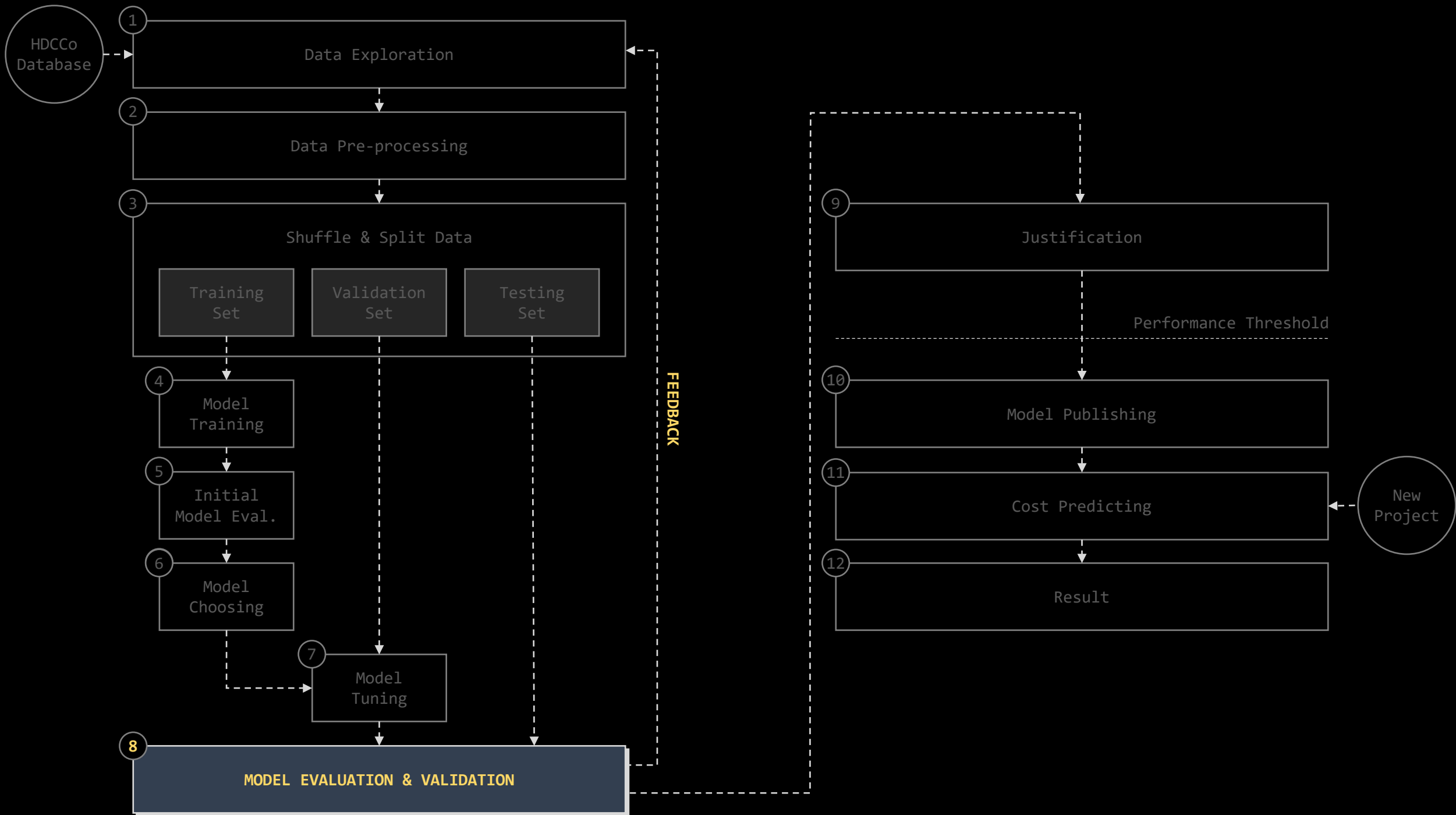
## RESULT

Model Evaluation & Validation  
Justification

## CONCLUSION

Reflection  
Improvement





		TRAIN	TEST
R2	• Ridge	0.90	0.66
	• RandomForestRegressor	0.94	0.65
	• SVR, linear	-0.18	-0.19
	• SVR, rbf	-0.18	-0.19
	• KNeighborsRegressor	0.68	0.60
	• XGBRegressor	0.99	0.78
	• Booster	0.84	0.60



R2

- Ridge
- RandomForestRegressor
- SVR, linear
- SVR, rbf
- KNeighborsRegressor
- **XGBRegressor**
- Booster

TRAIN

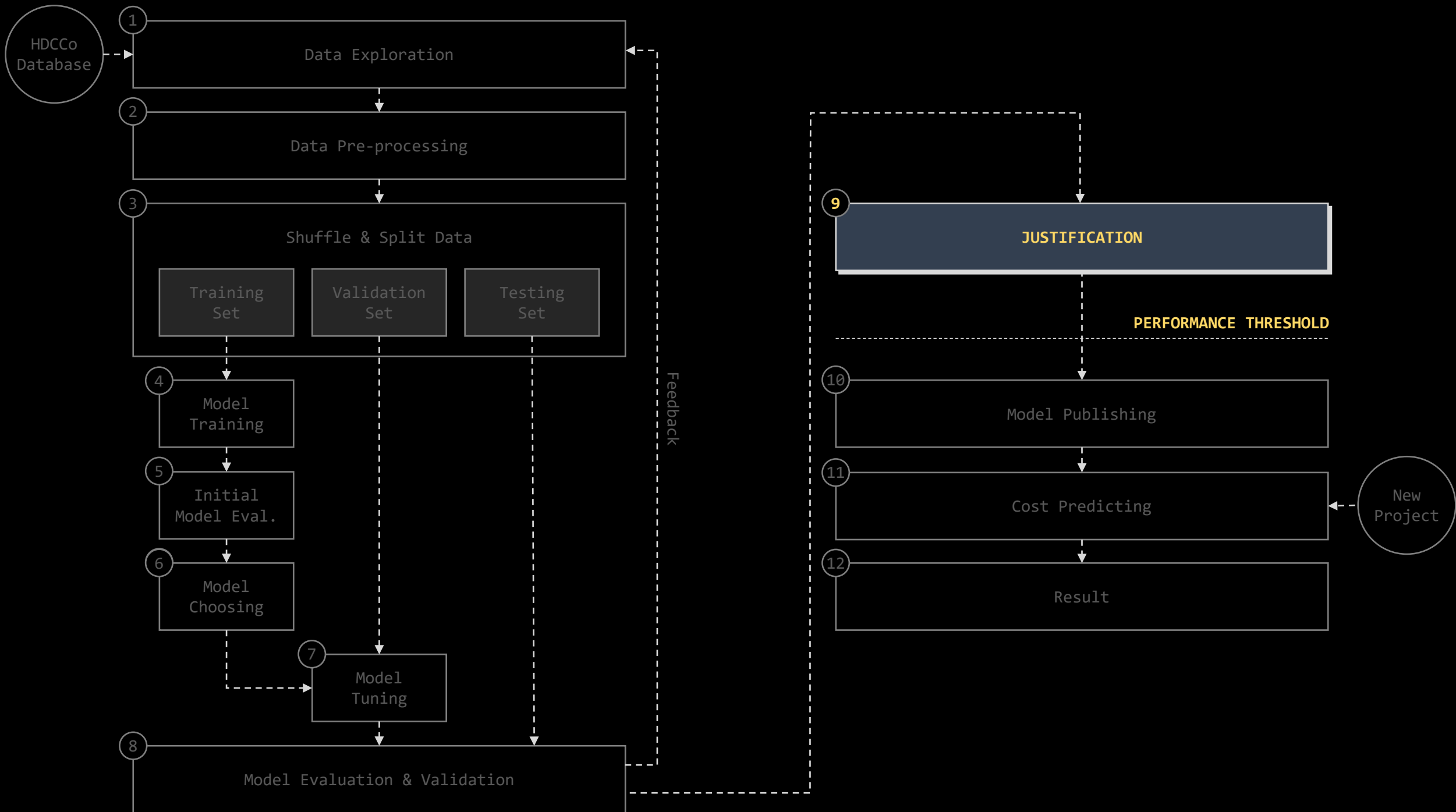
TEST

AFTER  
TUNING

0.90	0.66
0.94	0.65
-0.18	-0.19
-0.18	-0.19
0.68	0.60
0.99	<b>0.78</b>
0.84	0.60



0.93



Project Name	Primary Divisor	USF	Project Information in Detail	Amount	\$ / USF	USF	Bid Date	Duration	Trade Information in Detail	Max.	Min.	Avg.	CV	Number	Project List
Escalation	2019 - 4th Qtr.			\$ 2,602,625	\$ 407.47	6,387	3/8/2015	10.30	010110.Permits	\$ 4.14	\$ 0.06	\$ 2.05	81%	18	
Remove Extreme	Yes			\$ 5,587,421	\$ 174.93	31,941	8/22/2014	10.00	010120.Bonds	\$ 3.29	\$ 3.29	\$ 3.29	0%	1	
Client Type	4510_Software_and_Services			\$ 7,848,980	\$ 227.26	33,658	12/31/2014	16.85	011000.General_Conditions	\$ 42.51	\$ 1.91	\$ 11.14	81%	22	
Project Scope	Full TI Built-Out			\$ 40,039,969	\$ 278.63	144,741	9/11/2018	32.00	011510.Design_Consultants	\$ 4.74	\$ 0.06	\$ 1.80	116%	3	
Project Function	11.27_Office_Facility			\$ 719,997	\$ 150.63	4,783	3/21/2017	9.20	016120.Cleaning	\$ 0.74	\$ 0.74	\$ 0.74	0%	1	
Project Form				\$ 13,006,674	\$ 190.73	68,196	10/24/2012	12.00	019000.General_Requirements	\$ 29.67	\$ 0.13	\$ 6.43	93%	23	
Project Space				\$ 3,017,489	\$ 140.48	21,480	5/1/2014	10.00	019600.Testing_and_Inspections	\$ 3.97	\$ 0.07	\$ 0.79	179%	6	
Logistic Condition				\$ 21,207,607	\$ 141.73	149,631	6/30/2013	22.00	024100.Demolition	\$ 10.29	\$ 0.17	\$ 2.76	108%	17	
Design Method				\$ 2,854,330	\$ 105.17	27,139	5/25/2012	8.00	028200.Hazardous_Materials_Abatement	\$ 0.09	\$ 0.09	\$ 0.09	2%	2	
Fit & Finish				\$ 4,988,108	\$ 72.16	68,129	10/5/2012	19.00	033000.Cast-In-Place_Concrete	\$ 3.24	\$ 0.15	\$ 1.17	84%	11	
Architectural System				\$ 36,984,854	\$ 280.75	131,734	3/28/2017	29.00	033225.Concrete_Clean_and_Restore	\$ 2.00	\$ 0.70	\$ 1.20	47%	3	
MEPF System				\$ 863,194	\$ 68.79	12,548	3/10/2014	5.20	033800.Post-Tensioned_Concrete	\$ 1.52	\$ 1.52	\$ 1.52	0%	1	
Project Location				\$ 6,124,182	\$ 93.88	65,236	7/9/2017	19.40	042000.Unit_Masonry	\$ 0.09	\$ 0.09	\$ 0.09	0%	1	
Project Environment				\$ 39,352,021	\$ 629.09	62,554	10/10/2016	31.00	044000.Stone_Assemblies	\$ 0.01	\$ 0.01	\$ 0.01	0%	1	
LEED System				\$ 92,734,923	\$ 278.10	335,872	2/1/2016	49.00	051200.Structural_Steel_Framing	\$ 16.74	\$ 0.18	\$ 5.80	101%	6	
LEED Status				\$ 20,442,978	\$ 243.15	84,077	3/14/2012	29.00	055000.Metal_Fabrications	\$ 29.35	\$ 0.16	\$ 5.09	142%	16	
USF - Greater than				\$ 11,964,776	\$ 219.40	54,534	9/12/2014	20.00	061000.Rough_Carpentry	\$ 3.42	\$ 0.01	\$ 0.81	106%	17	
USF - Less than				\$ 9,989,070	\$ 134.36	74,198	8/1/2011	15.57	064000.Architectural_Woodwork	\$ 57.48	\$ 0.89	\$ 15.15	97%	23	
Bid Date - Later than				\$ 34,788,441	\$ 277.63	125,348	11/18/2013	80.00	071000.Dampproofing_and_Waterproofing	\$ 1.51	\$ 0.14	\$ 0.46	114%	5	
Bid Date - Earlier than				\$ 596,827	\$ 269.05	2,226	7/29/2017	10.00	076000.Membrane_Roofing	\$ 0.06	\$ 0.06	\$ 0.06	0%	1	
Number of Projects	25			\$ 122,339,320	\$ 239.41	511,014	3/5/2018	113.00	078100.Applied_Fireproofing	\$ 0.56	\$ 0.03	\$ 0.20	90%	6	
Amount - Maximum	\$	122,339,320		\$ 21,274,630	\$ 382.24	55,658	12/1/2017	30.20	081000.Doors_and_Frames	\$ 14.19	\$ 2.03	\$ 5.31	62%	23	
Amount - Minimum	\$	596,827		\$ 8,681,453	\$ 472.82	18,357	12/1/2017	27.40	083000.Specialty_Doors_and_Frames	\$ 2.81	\$ 0.06	\$ 1.07	80%	11	
Amount - Average	\$	23,519,800		\$ 41,062,544	\$ 335.33	122,454	4/9/2014	38.20	083300.Colling_Doors_and_Grilles	\$ 0.62	\$ 0.62	\$ 0.62	0%	1	
Amount - Coeff. of Variation		125%		\$ 38,140,586	\$ 258.74	151,275	4/22/2015	32.00	088000.Interior_Glazing	\$ 12.88	\$ 0.78	\$ 4.96	58%	23	
\$ / USF - Maximum	\$	629.09							092900.Gypsum_Board_Assemblies	\$ 43.30	\$ 3.89	\$ 17.80	54%	23	
\$ / USF - Minimum	\$	68.79							093000.Tiling	\$ 6.96	\$ 0.06	\$ 2.09	83%	20	
\$ / USF - Average	\$	242.71							095000.Ceilings	\$ 47.47	\$ 1.84	\$ 10.26	87%	22	
\$ / USF - Coeff. of Variation		54%							098000.Flooring	\$ 14.80	\$ 2.59	\$ 7.35	38%	22	
USF - Maximum	511,014								098700.Fluid-Applied_Flooring	\$ 7.00	\$ 0.54	\$ 3.02	94%	3	
USF - Minimum	2,226								098800.Access_Flooring	\$ 59.49	\$ 0.49	\$ 21.66	77%	7	
USF - Average	94,567								097000.Graphics	\$ 1.90	\$ 0.65	\$ 1.38	38%	3	
USF - Coeff. of Variation	119%								099000.Painting_and_Coating	\$ 25.20	\$ 1.13	\$ 6.26	79%	22	
Duration - Maximum	113.00								101100.Visual_Display_Surfaces	\$ 1.96	\$ 0.08	\$ 1.29	67%	3	
Duration - Minimum	5.20								101400.Signage	\$ 3.30	\$ 0.10	\$ 1.41	92%	8	
Duration - Average	27.13								102000.Interior_Specialties	\$ 5.56	\$ 0.13	\$ 1.67	92%	14	
Duration - Coeff. of Variation	89%								102113.Toilet_Compartments_and_Accessories	\$ 8.01	\$ 0.08	\$ 1.23	158%	14	
									102219.Demountable_Partitions	\$ 0.81	\$ 0.81	\$ 0.81	0%	1	
									102228.Operable_Partitions	\$ 1.14	\$ 0.55	\$ 0.84	35%	2	
									104400.Fire_Protection_Specialties	\$ 0.17	\$ 0.01	\$ 0.09	77%	3	
									111100.Parking_and>Loading_Dock_Equipment	\$ 0.03	\$ 0.03	\$ 0.03	0%	1	
									113000.Appliances	\$ 6.59	\$ 0.02	\$ 1.34	140%	13	
									114000.Foodservice_Equipment	\$ 24.78	\$ 0.98	\$ 9.89	83%	7	
									116100.Theater_and_Stage_Equipment	\$ 3.07	\$ 3.07	\$ 3.07	0%	1	
									122000.Window_Treatments	\$ 10.20	\$ 0.18	\$ 2.42	103%	22	
									126000.Furniture	\$ 2.89	\$ 0.11	\$ 1.50	92%	2	
									142000.Elevators	\$ 21.61	\$ 0.08	\$ 3.36	195%	9	
									148000.Scaffolding	\$ 0.27	\$ 0.11	\$ 0.19	43%	2	
									211000.Fire_Sprinklers	\$ 7.66	\$ 0.92	\$ 3.59	49%	23	

Project Name	Primary Divider	USF	Project Information in Detail	Amount	\$ / USF	USF	27.13	Trade Information in Detail	Max.	Min.	Avg.	CV.	Project	Number	List
Escalation	2019 - 4th Qtr.			\$ 2,602,626	\$ 407.47	6,387	3/8/2015	10.30	\$ 4.14	\$ 0.06	\$ 2.05	81%		18	
Remove Extreme	Yes			\$ 5,587,421	\$ 174.83	31,941	6/22/2014	10.00	\$ 3.29	\$ 3.29	\$ 3.29	0%		1	
Client Type	4510_Software_and_Services			\$ 7,846,980	\$ 227.26	33,658	12/31/2014	16.85	\$ 42.51	\$ 1.91	\$ 11.14	81%		22	
Project Scope	Full TI Built-Out			\$ 40,039,969	\$ 278.63	144,741	9/11/2018	32.00	\$ 4.74	\$ 0.06	\$ 1.80	116%		3	
Project Function	11.27_Office_Facility			\$ 719,997	\$ 150.63	4,783	3/21/2017	9.20	\$ 0.74	\$ 0.74	\$ 0.74	0%		1	
Project Form				\$ 13,006,674	\$ 190.73	68,196	10/24/2012	12.00	\$ 29.87	\$ 0.13	\$ 6.43	93%		23	
Project Space				\$ 3,017,489	\$ 140.48	21,480	5/1/2014	10.00	\$ 3.97	\$ 0.07	\$ 0.79	179%		6	
Logistic Condition				\$ 21,207,607	\$ 141.73	149,631	6/30/2013	22.00	\$ 10.29	\$ 0.17	\$ 2.76	108%		17	
Design Method				\$ 2,854,330	\$ 105.17	27,139	5/25/2012	8.00	\$ 0.09	\$ 0.09	\$ 0.09	2%		2	
Fit & Finish				\$ 4,988,108	\$ 72.16	69,129	10/5/2012	19.00	\$ 3.24	\$ 0.15	\$ 1.17	84%		11	
Architectural System				\$ 36,984,854	\$ 280.75	131,734	3/28/2017	29.00	\$ 2.00	\$ 0.70	\$ 1.20	47%		3	
MEPF System				\$ 863,194	\$ 68.79	12,548	3/10/2014	5.20	\$ 1.52	\$ 1.52	\$ 1.52	0%		1	
Project Location				\$ 6,124,182	\$ 93.88	65,236	7/9/2017	19.40	\$ 0.09	\$ 0.09	\$ 0.09	0%		1	
Project Environment				\$ 39,352,021	\$ 629.09	62,554	10/10/2016	31.00	\$ 0.01	\$ 0.01	\$ 0.01	0%		1	
LEED System				\$ 92,734,923	\$ 278.10	335,872	2/1/2016	49.00	\$ 16.74	\$ 0.18	\$ 5.80	101%		6	
LEED Status				\$ 20,442,978	\$ 243.15	84,077	3/14/2012	29.00	\$ 29.35	\$ 0.16	\$ 5.09	142%		16	
USF - Greater than				\$ 11,984,776	\$ 219.40	54,534	9/12/2014	20.00	\$ 3.42	\$ 0.01	\$ 0.81	106%		17	
USF - Less than				\$ 9,989,070	\$ 134.36	74,198	8/1/2011	15.57	\$ 57.48	\$ 0.89	\$ 15.15	97%		23	
Bid Date - Later than				\$ 34,788,441	\$ 277.63	125,348	11/18/2013	80.00	\$ 1.81	\$ 0.14	\$ 0.46	114%		5	
Bid Date - Earlier than				\$ 598,827	\$ 269.05	2,226	7/29/2011	16.70	\$ 0.06	\$ 0.10	\$ 0.08	0%		1	
Number of Projects	28			\$ 122,339,320	\$ 239.41	511,014	3/5/2011	10.00	\$ 0.56	\$ 0.10	\$ 0.20	90%		6	
Amount - Maximum	\$ 122,339,320			\$ 21,274,630	\$ 382.24	55,858	12/1/2017	21.20	\$ 14.19	\$ 2.03	\$ 1.93	92%		23	
Amount - Minimum	\$ 598,827			\$ 8,681,453	\$ 472.82	18,357	12/1/2017	10.00	\$ 2.81	\$ 0.10	\$ 1.10	0%		11	
Amount - Average	\$ 23,519,800			\$ 41,062,544	\$ 335.33	122,454	4/9/2014	38.20	\$ 0.62	\$ 0.62	\$ 0.62	0%		1	
Amount - Coeff. of Variation	125%			\$ 39,140,586	\$ 258.74	151,275	4/22/2015	32.00	\$ 12.88	\$ 0.78	\$ 4.96	58%		23	
\$ / USF - Maximum	\$ 629.09								\$ 43.30	\$ 3.89	\$ 1.70	54%		23	
\$ / USF - Minimum	\$ 68.79								\$ 6.96	\$ 0.06	\$ 2.09	83%		20	
\$ / USF - Average	\$ 242.71								\$ 47.47	\$ 1.84	\$ 10.26	87%		22	
\$ / USF - Coeff. of Variation	54%								\$ 14.80	\$ 2.59	\$ 7.35	38%		22	
USF - Maximum	511,014								\$ 7.00	\$ 0.54	\$ 3.02	94%		3	
USF - Minimum	2,226								\$ 69.700	\$ 0.49	\$ 21.68	77%		7	
USF - Average	94,567								\$ 1.90	\$ 0.65	\$ 1.38	38%		3	
USF - Coeff. of Variation	119%								\$ 69.000	\$ 1.13	\$ 8.26	79%		22	
Duration - Maximum	113.00								\$ 101100	\$ 0.08	\$ 1.29	87%		3	
Duration - Minimum	5.20								\$ 101400	\$ 0.10	\$ 1.41	92%		8	
Duration - Average	27.13								\$ 102000	\$ 0.13	\$ 1.67	92%		14	
Duration - Coeff. of Variation	89%								\$ 102113	\$ 0.08	\$ 1.23	158%		14	
									\$ 102219	\$ 0.81	\$ 0.81	0%		1	
									\$ 102228	\$ 1.14	\$ 0.65	0.84	35%	2	
									\$ 104400	\$ 0.17	\$ 0.01	0.09	77%	3	
									\$ 111100	\$ 0.03	\$ 0.03	0.03	0%	1	
									\$ 113000	\$ 6.59	\$ 0.02	1.34	140%	13	
									\$ 114000	\$ 24.78	\$ 0.98	9.89	83%	7	
									\$ 116100	\$ 3.07	\$ 3.07	3.07	0%	1	
									\$ 122000	\$ 10.20	\$ 0.18	2.42	103%	22	
									\$ 125000	\$ 2.89	\$ 0.11	1.50	92%	2	
									\$ 142000	\$ 21.61	\$ 0.08	3.38	196%	9	
									\$ 148000	\$ 0.27	\$ 0.11	0.19	43%	2	
									\$ 211000	\$ 7.68	\$ 0.82	3.59	49%	23	

CLIENT  
TYPE

0.79

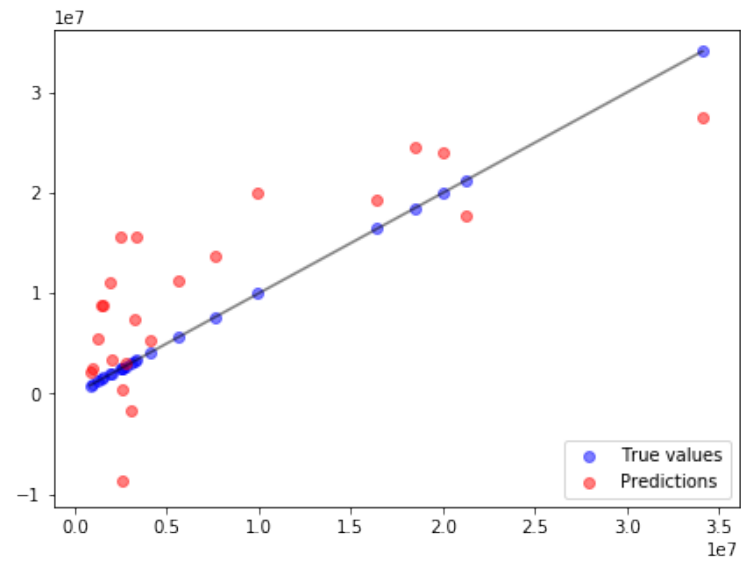
PROJECT  
SCOPE

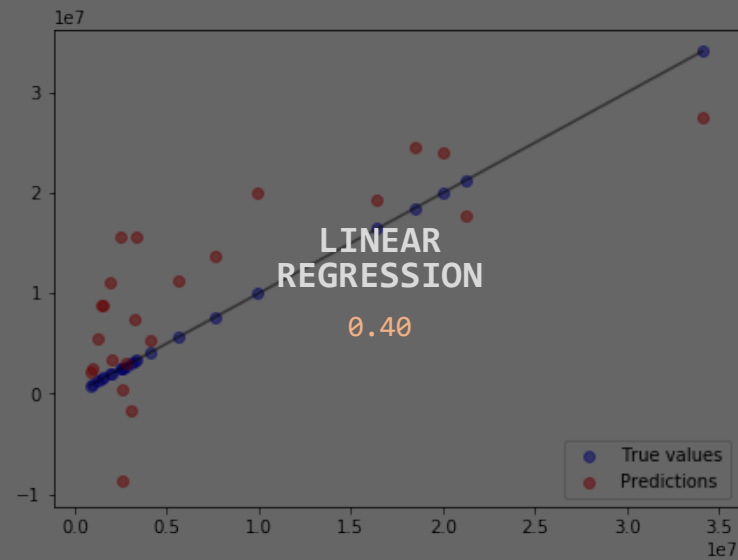
0.86

PROJECT  
FUNCTION

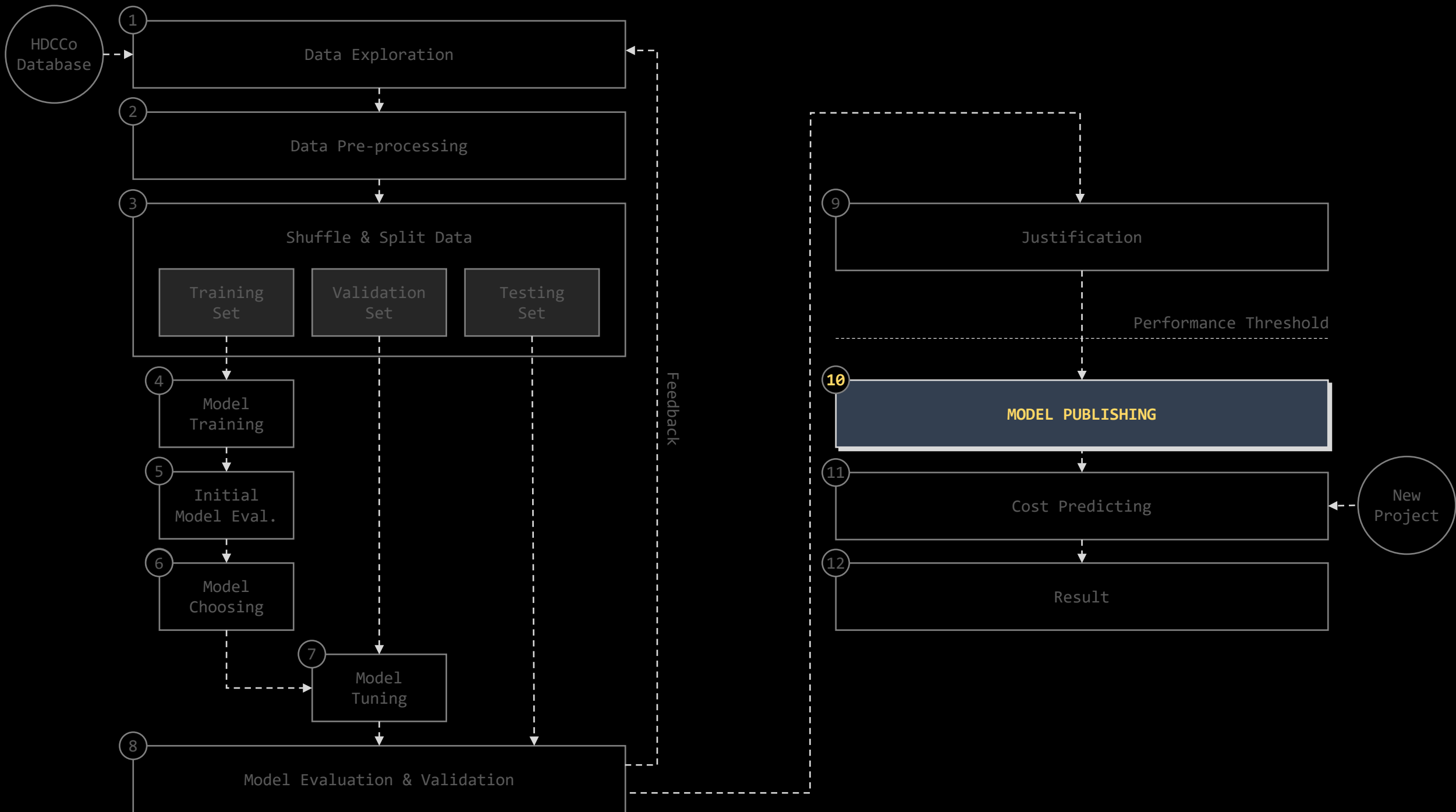
0.80

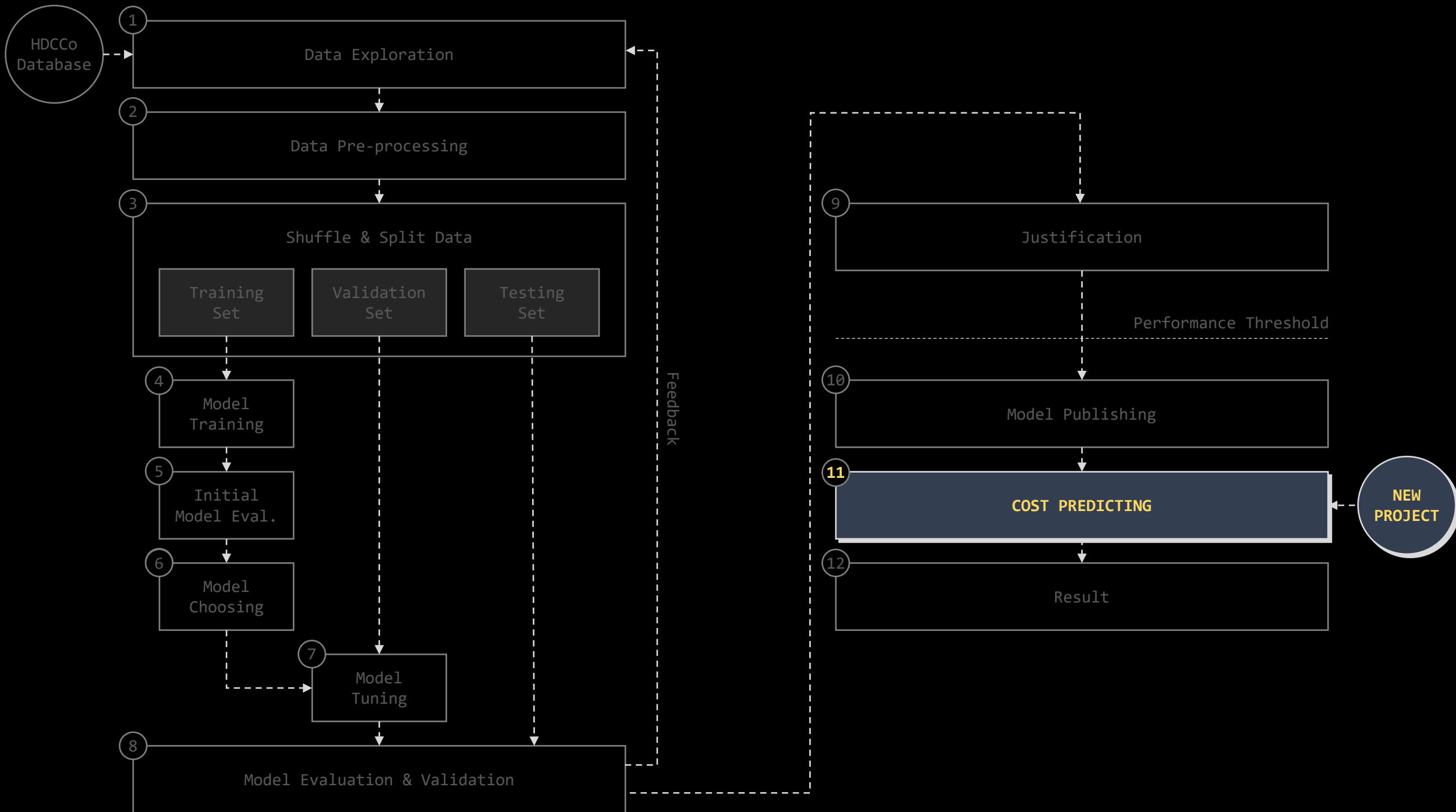
Sensitive  
Data

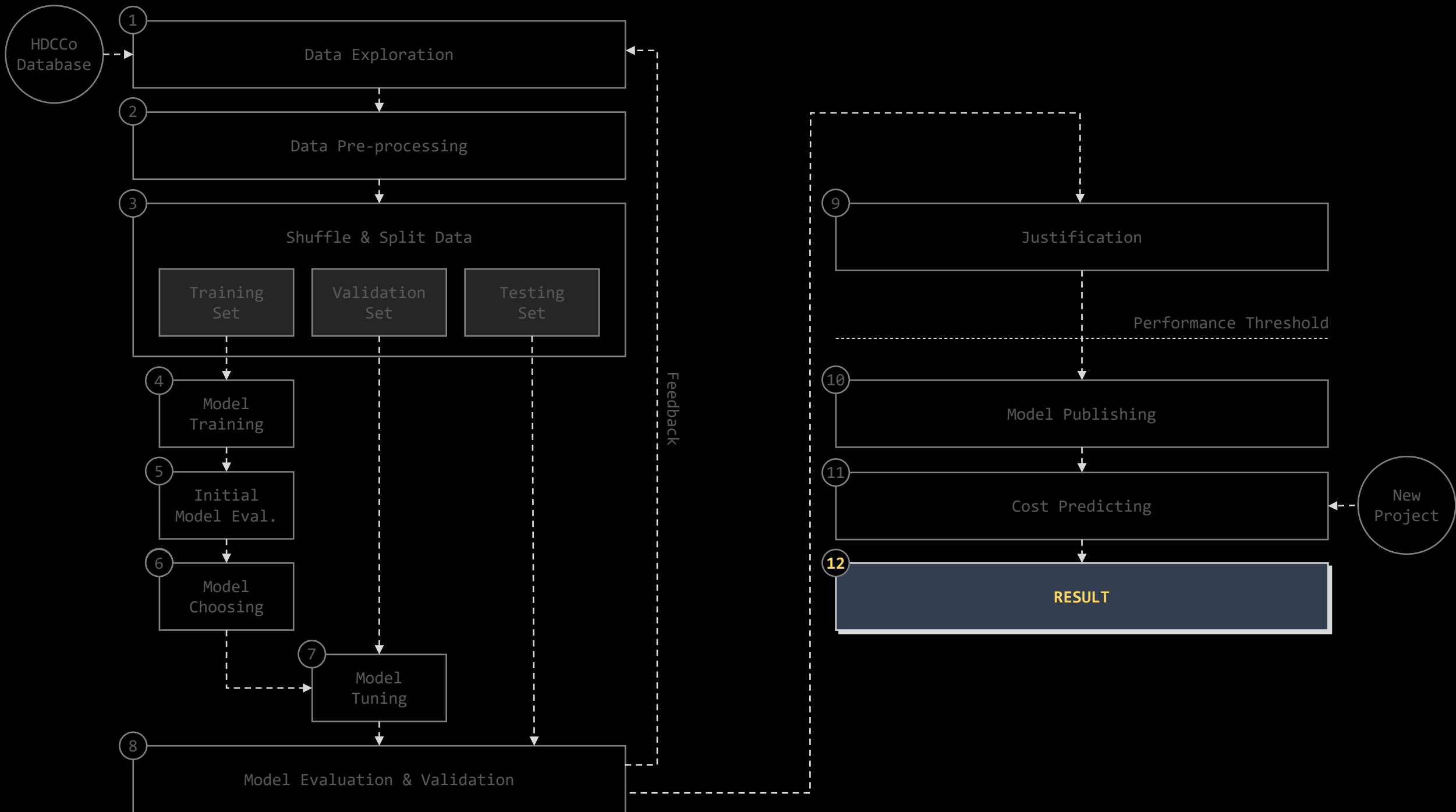


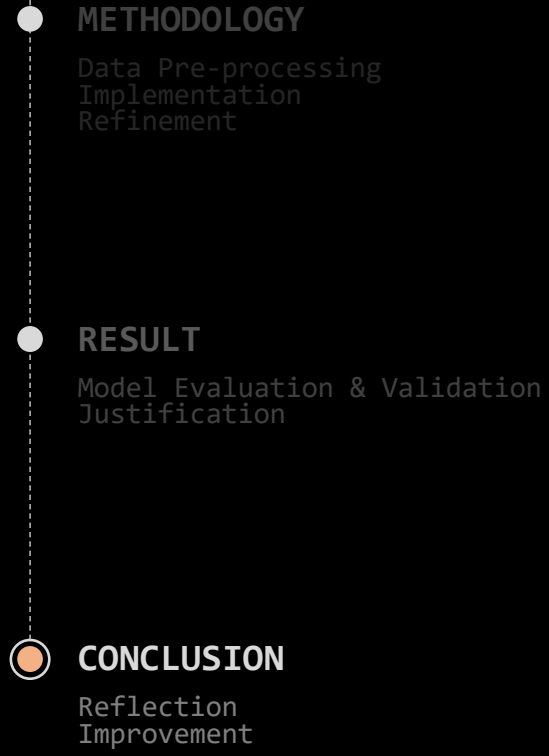






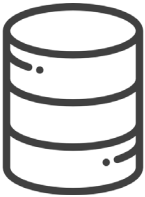


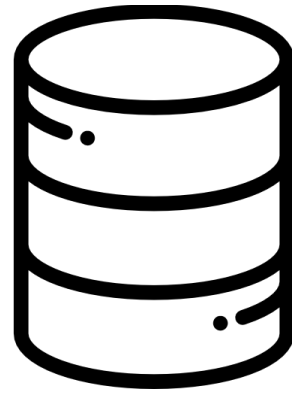




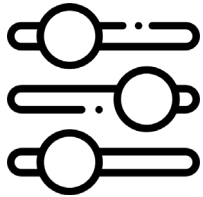












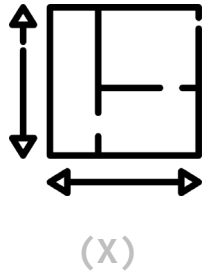
(X)

#### NUMERICAL INPUT

- Usable square footage (USF)
- Rentable square footage (RSF)
- Number of floors
- Construction duration

#### NON-NUMERICAL INPUT

- Client type
- Project scope
- Project function
- Project form
- Project location
- Project environment
- Logistic condition
- Design method
- Fit & finish
- Architecture system



#### NUMERICAL INPUT

- Room Count
- Room Area
- Room Perimeter

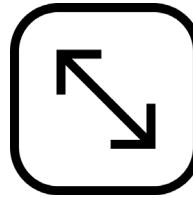
#### NON-NUMERICAL INPUT

- Client type
- Project scope
- Project function
- Project form
- Project location
- Project environment
- Logistic condition
- Design method
- Fit & finish
- Architecture system



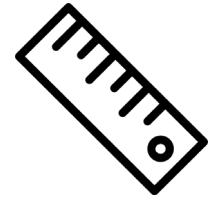
### ROOM COUNT

Door  
Equipment  
Specialty  
Fixture



### ROOM AREA

Flooring  
Ceiling  
Fire sprinkler  
HVAC



### ROOM PERIMETER

Drywall  
Glazing  
Wall finish  
Window treatment



CONSTRUCTION COST



CONSTRUCTION COST



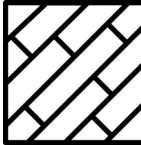
ARCH. WOODWORK

064000



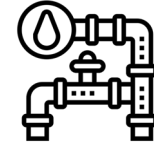
FIRE SPRINKLER

211000



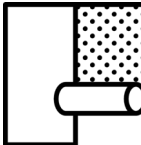
FLOORING

096000



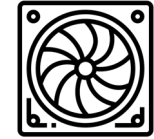
PLUMBING

221000



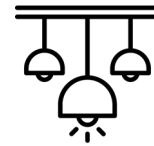
PAINTING & COATING

099000



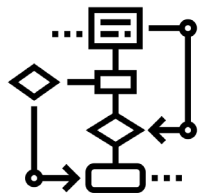
HVAC

231000



ELECTRICAL

261000



## ML-RELATED

### TRIAL-AND-ERROR

- Data preprocessing
- Algorithms and techniques
- Free-form visualization
- This is what makes this a science!

### FEATURE SELECTION

- feature\_selection
- SelectKBest
- SelectPercentile
- Likelihood encoding of categorical features
- Smoothing



## DEPLOYMENT

### PLATFORM INTEGRATION

- AWS/Azure
- Data studio/Power BI/Tableau

### USER INTERFACE

- Web browser
- Autodesk Forge



LIUS@HDCCO.COM