TCS Technical Assessment

Prudential BMI Case Study
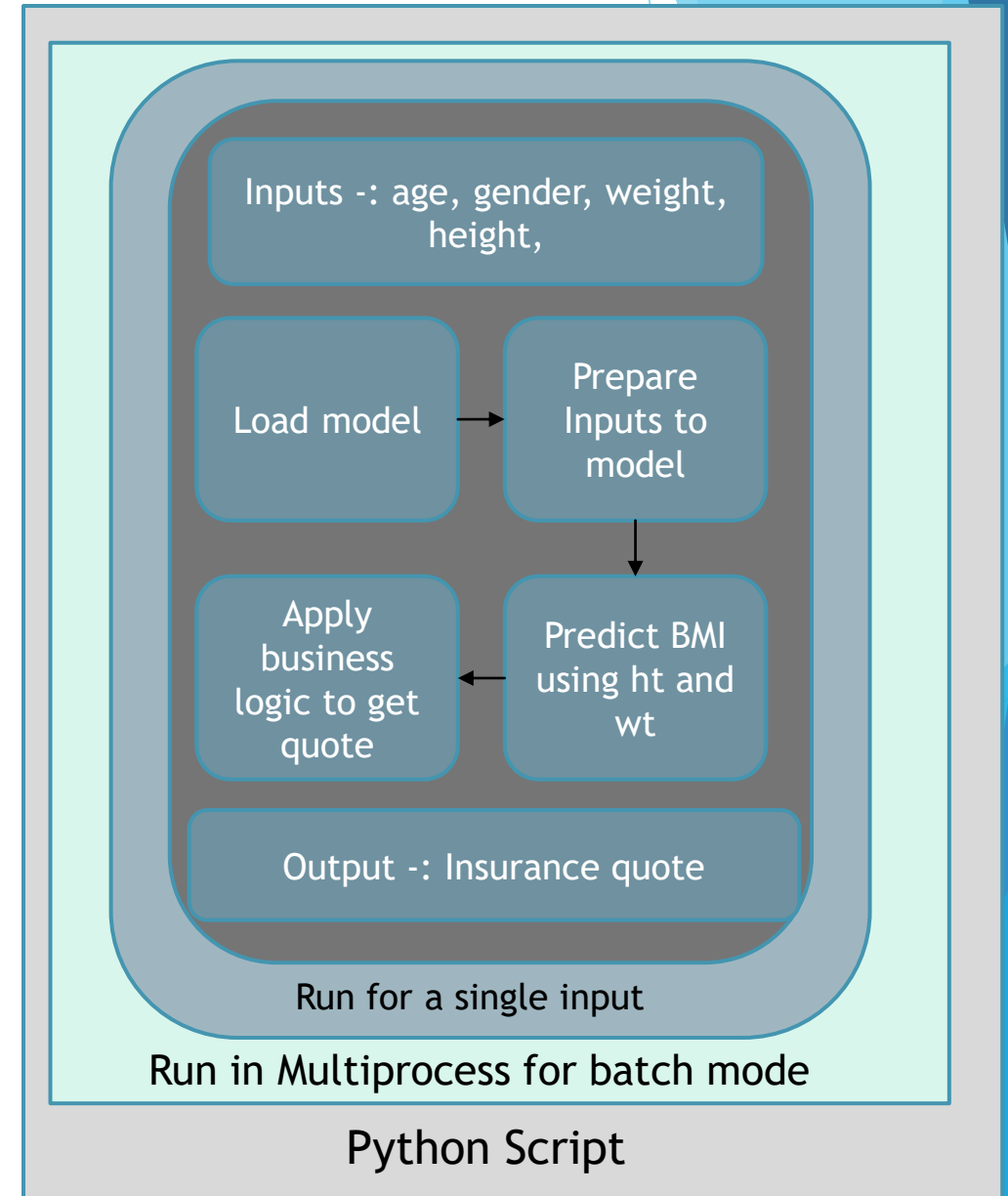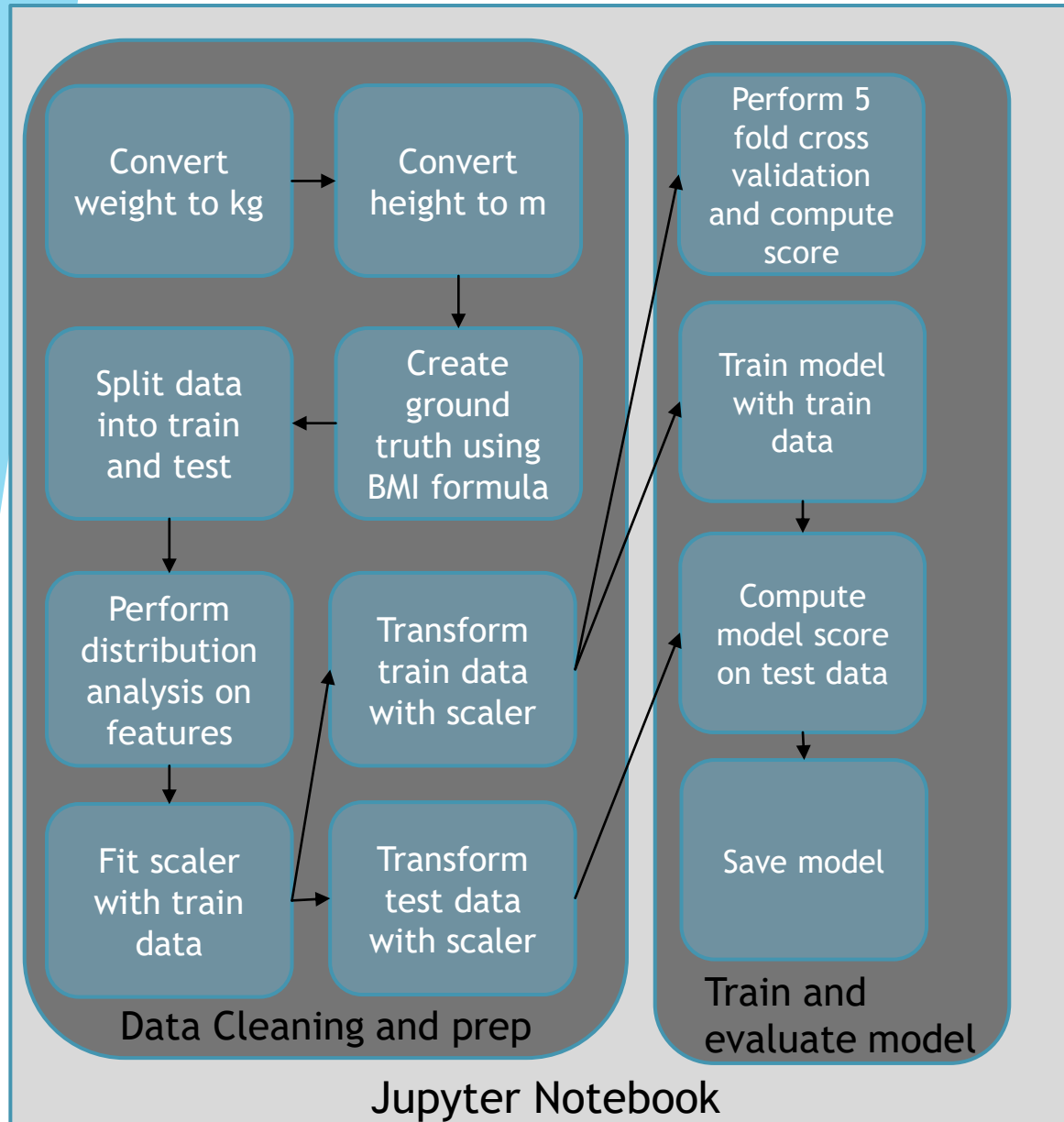
Krishna Hariramani
20/12/2021

Agenda

- High Level Overview
- Model Building and experimentation
  - Data Cleaning & Preparation
  - Training Model and Evaluation
- Operational the Model

High Level Overview

Jupyter Notebook

Data Cleaning and prep

- Convert weight to kg
- Convert height to m
- Create ground truth using BMI formula
- Split data into train and test
- Perform distribution analysis on features
- Transform train data with scaler
- Fit scaler with train data
- Transform test data with scaler

Train and evaluate model

- Perform 5 fold cross validation and compute score
- Train model with train data
- Compute model score on test data
- Save model

Python Script

Run in Multiprocess for batch mode

Run for a single input

- Inputs -: age, gender, weight, height,
- Load model
- Prepare Inputs to model
- Apply business logic to get quote
- Predict BMI using ht and wt
- Output -: Insurance quote

Data Cleaning and Preparation – Converting weight and height

	AppID	Ins_Age	Ins_Gender	Ht	Wt	IssueDate
0	56372	31	Male	510	185	NaN
1	34565	35	Male	510	205	NaN
2	57732	45	Female	510	125	NaN
3	87324	38	Male	503	175	NaN
4	12323	39	Female	600	252	NaN

	AppID	Ins_Age	Ins_Gender	Ht	Wt	IssueDate	Wt_kg
0	56372	31	Male	510	185	NaN	83.914520
1	34565	35	Male	510	205	NaN	92.986360
2	57732	45	Female	510	125	NaN	56.699000
3	87324	38	Male	503	175	NaN	79.378600
4	12323	39	Female	600	252	NaN	114.305184

	AppID	Ins_Age	Ins_Gender	Ht	Wt	IssueDate	Wt_kg	Ht_modified	Ht_ft	Ht_in	Ht_m
0	56372	31	Male	510	185	NaN	83.914520	5.1	5	1	1.5494
1	34565	35	Male	510	205	NaN	92.986360	5.1	5	1	1.5494
2	57732	45	Female	510	125	NaN	56.699000	5.1	5	1	1.5494
3	87324	38	Male	503	175	NaN	79.378600	5.03	5	3	1.6002
4	12323	39	Female	600	252	NaN	114.305184	6.0	6	0	1.8288

Data Cleaning and Preparation – Calculate ground truth and Split data

	Ins_Age	Ins_Gender	Wt_kg	Ht_m	BMI
0	31	Male	83.914520	1.5494	34.955057
1	35	Male	92.986360	1.5494	38.733982
2	45	Female	56.699000	1.5494	23.618282
3	38	Male	79.378600	1.6002	30.999515
4	39	Female	114.305184	1.8288	34.176966

(85, 5)

	Ins_Age	Ins_Gender	Wt_kg	Ht_m	BMI
42	19	Female	81.64656	1.8034	25.104626
77	36	Female	77.11064	1.7526	25.104333
38	46	Female	54.43104	1.6764	19.368292
56	24	Male	81.64656	1.9304	21.910045
75	47	Male	81.64656	1.7018	28.191673

(15, 5)

	Ins_Age	Ins_Gender	Wt_kg	Ht_m	BMI
8	40	Female	70.30676	1.8288	21.021546
70	37	Male	72.57472	1.6764	25.824390
82	36	Male	63.50288	1.5748	25.606051
28	19	Female	77.11064	1.8542	22.428548
63	21	Male	72.57472	1.6002	28.342414

Data Cleaning and Preparation – Fit scaler on train and transform test

Out[10]:

	Ins_Age	Ins_Gender	Wt_kg	Ht_m	BMI
42	19	Female	0.576398	0.879392	25.104626
77	36	Female	0.205794	0.440985	25.104333
38	46	Female	-1.647224	-0.216624	19.368292
56	24	Male	0.576398	1.975408	21.910045
75	47	Male	0.576398	0.002579	28.191673

In [11]: `test.head()`

Out[11]:

	Ins_Age	Ins_Gender	Wt_kg	Ht_m	BMI
8	40	Female	-0.350111	1.098595	21.021546
70	37	Male	-0.164810	-0.216624	25.824390
82	36	Male	-0.906017	-1.093437	25.606051
28	19	Female	0.205794	1.317798	22.428548
63	21	Male	-0.164810	-0.874234	28.342414

Data Cleaning and Preparation – Feature Distributions

➢ **Weight -:**
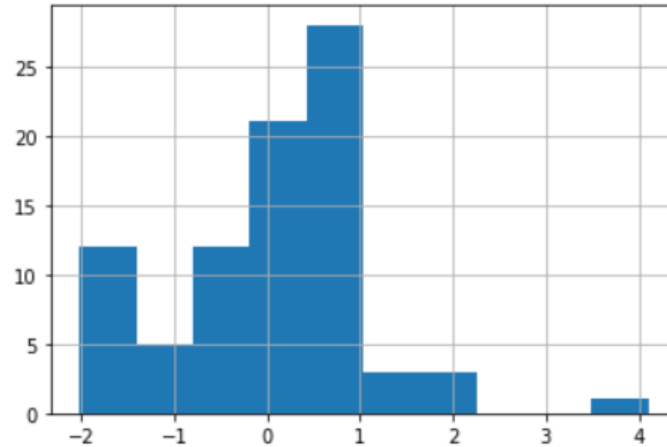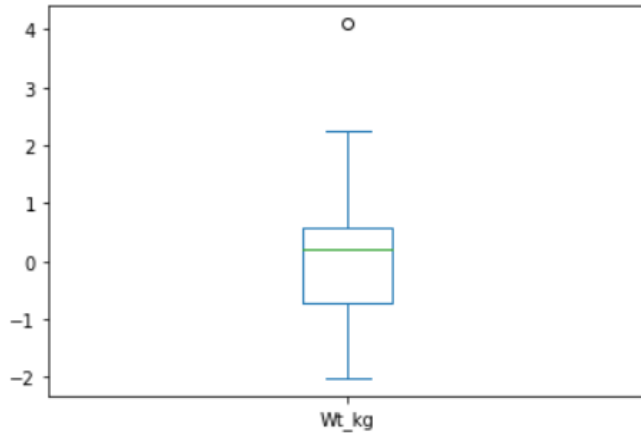
➢ **Height -:**
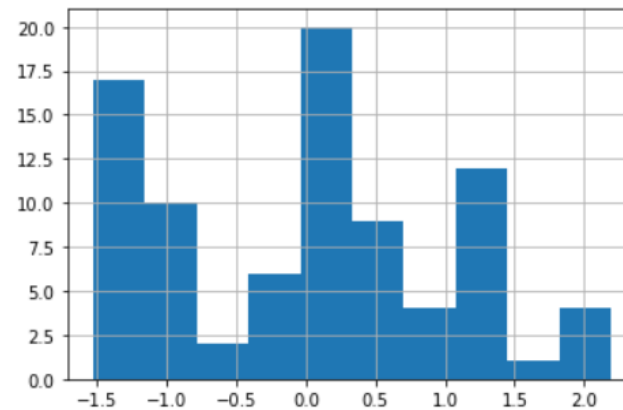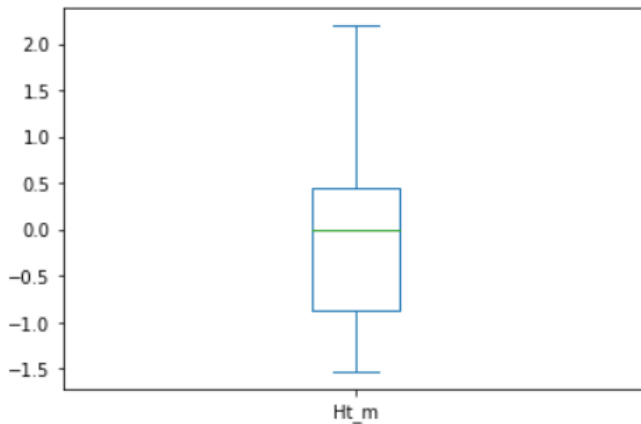
Train model and evaluation

➢ **Linear Regression -:**

```
R2 across 5 score arr [0.98736258 0.97255642 0.98273183 0.98547715 0.9708301 ]
Avg R2 for cross validation: 0.979791618808078
R2 for test set 0.96820093242257
```

➢ **Support Vector Regressor -:**

```
R2 across 5 score arr [0.96710477 0.19853876 0.87052928 0.95072015 0.8721909 ]
Avg R2: 0.7718167720548602
R2 for test set 0.8143370757933495
```

Operationalising the model