Auto ML – Model to select a Model

Jagadeesha (163059002),
Ebin Chacko (153079009),
Nagma Khan (153079030),
Sanjeev Kumar (16305R008),
Vivek Mishra (143079005)

Why?

- Unmet need of skilled data analysts
- Requirement of algorithms which decide a model for a given classification/regression task
- Also tunes its hyperparameters and does optimal feature reduction
- Solution to all this is AUTO-ML

The REALITY

Hyper-parameter tuning

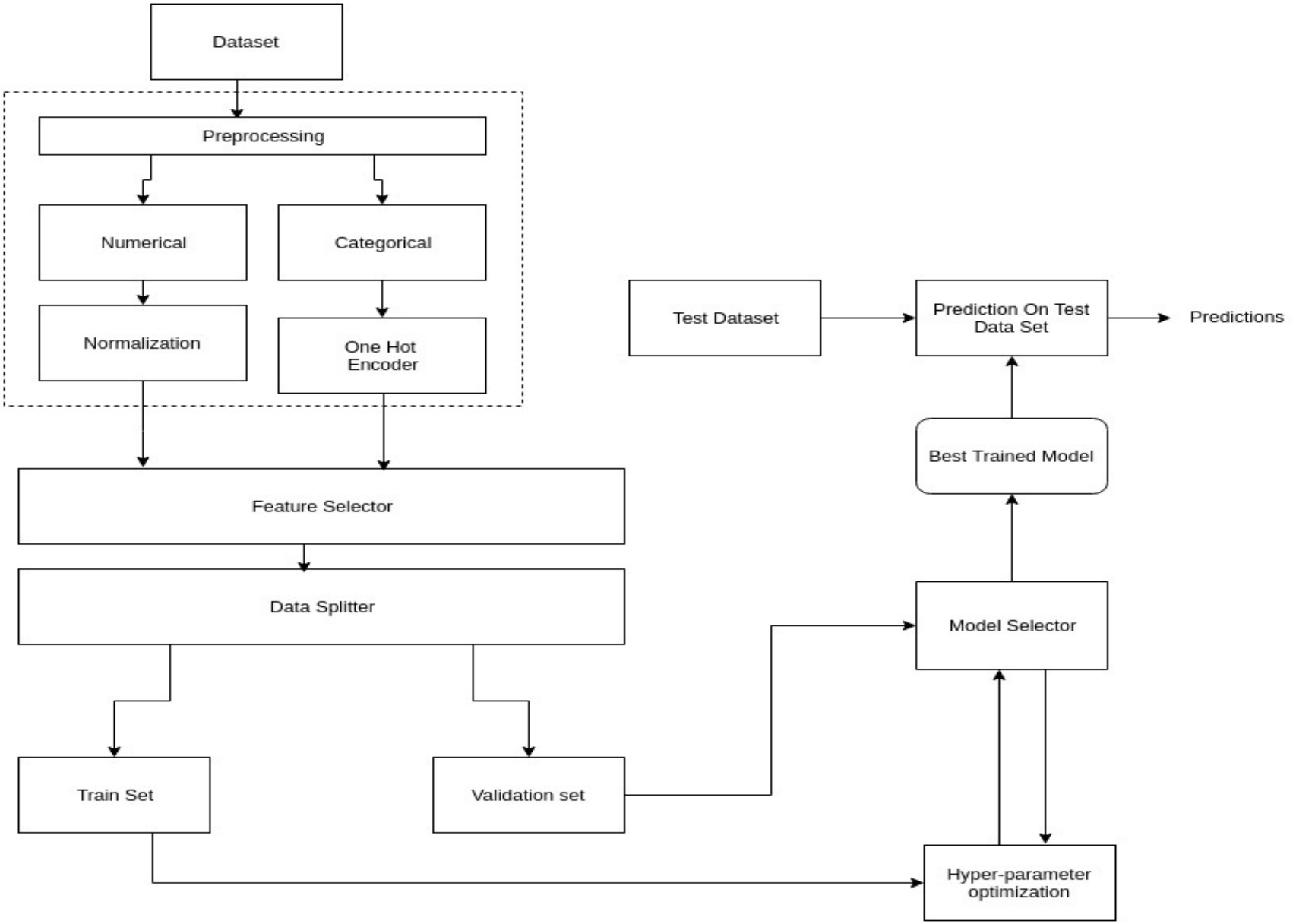
Trained model

TRAINING DATA

Answer y

Query x

Model Framework



Dataset Used

Online News Popularity Dataset

- Source: UCI machine learning repository
- Data Description: This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity).
- Number of Attributes: 61
- Number of Train Data Samples used: 39,644
- Number of Test Data Samples used: 9000

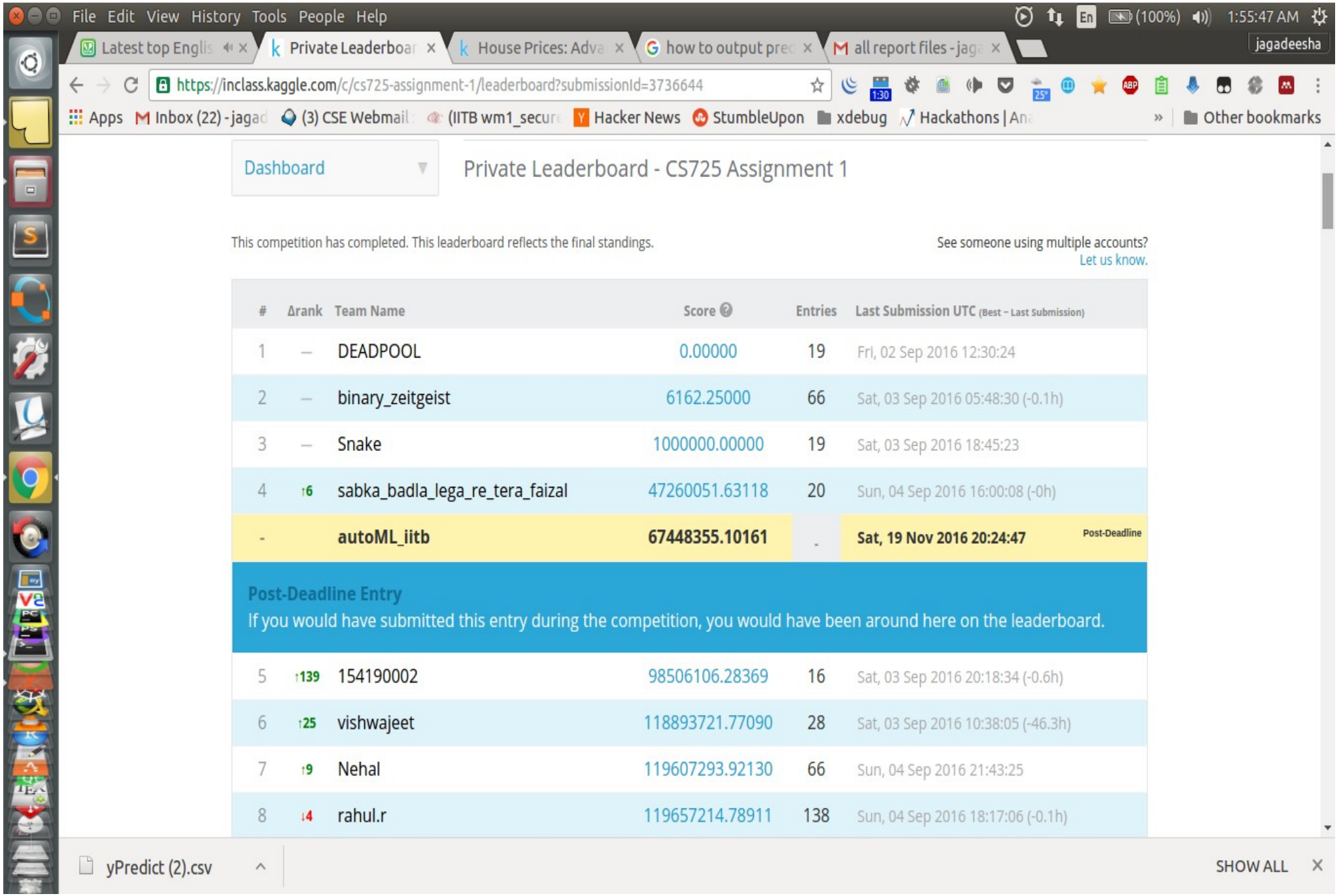
Dataset Used

House Prices Prediction

- Source: Kaggle
- Data Description: This playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.
- Number of Attributes: 80
- Number of Train Data Samples used: 1400
- Number of Test Data Samples used: 1400

Results

- Best Model : Support Vector Regression with RBF Kernel
- Pre-processing: Feature Reduction with Lasso
- Optimal Parameters: C=216.5 and gamma = 0.3753
- Error on Test Dataset (Mean Absolute Error): 2252.5



File Edit View History Tools People Help

Latest top Englis | Private Leaderboar | House Prices: Adva | how to output prec | all report files - jaga

jagadeesha

https://inclass.kaggle.com/c/cs725-assignment-1/leaderboard?submissionId=3736644

Apps | Inbox (22) - jagad | (3) CSE Webmail | (IITB wm1_secure | Hacker News | StumbleUpon | xdebug | Hackathons | Ana | Other bookmarks

Dashboard

Private Leaderboard - CS725 Assignment 1

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts?
Let us know.

#	Δrank	Team Name	Score	Entries	Last Submission UTC (Best – Last Submission)
1	–	DEADPOOL	0.00000	19	Fri, 02 Sep 2016 12:30:24
2	–	binary_zeitgeist	6162.25000	66	Sat, 03 Sep 2016 05:48:30 (-0.1h)
3	–	Snake	1000000.00000	19	Sat, 03 Sep 2016 18:45:23
4	↑6	sabka_badla_lega_re_tera_faizal	47260051.63118	20	Sun, 04 Sep 2016 16:00:08 (-0h)
-		**autoML_iitb**	**67448355.10161**	-	**Sat, 19 Nov 2016 20:24:47** Post-Deadline

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

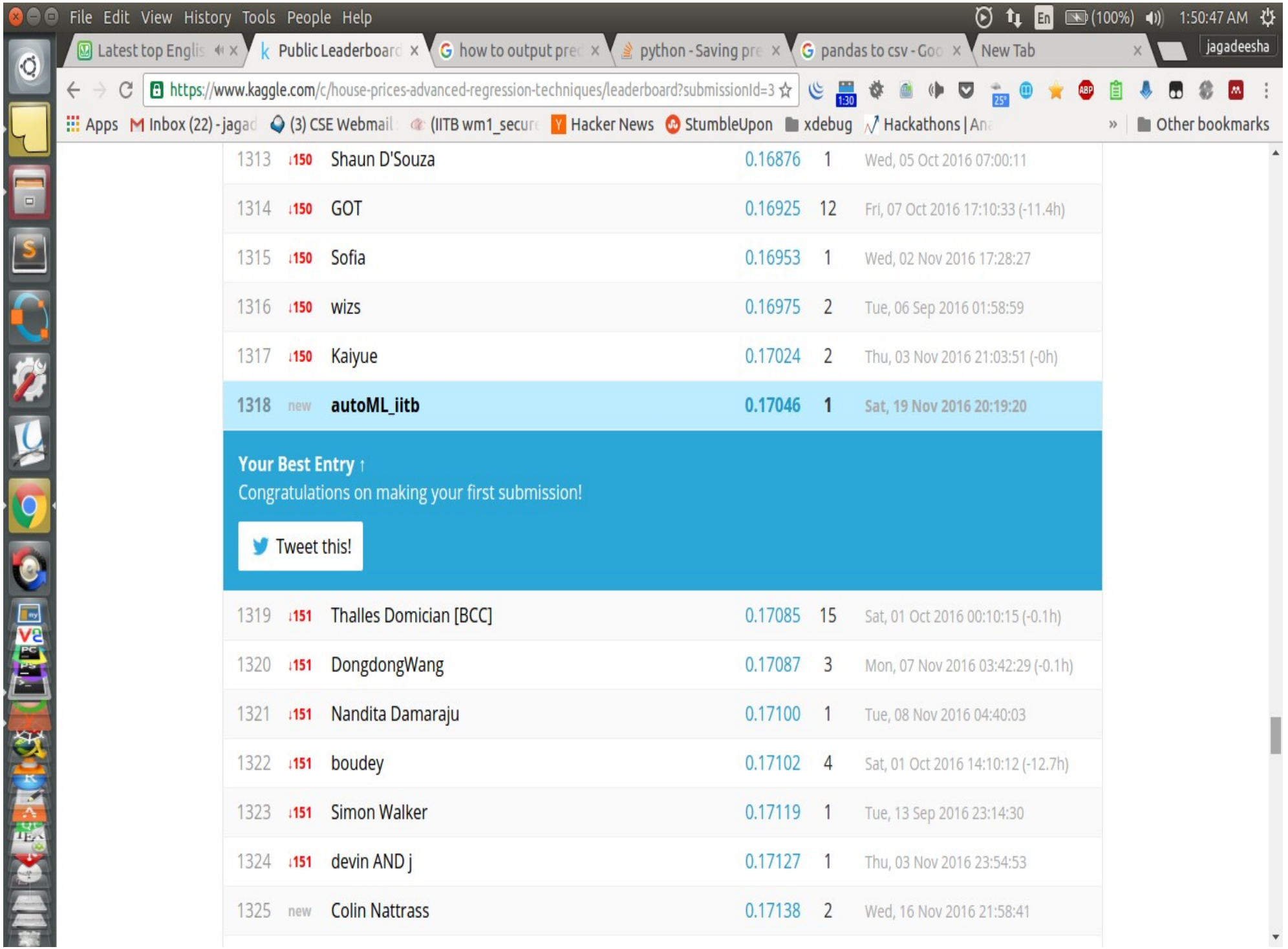
#	Δrank	Team Name	Score	Entries	Last Submission UTC (Best – Last Submission)
5	↑139	154190002	98506106.28369	16	Sat, 03 Sep 2016 20:18:34 (-0.6h)
6	↑25	vishwajeet	118893721.77090	28	Sat, 03 Sep 2016 10:38:05 (-46.3h)
7	↑9	Nehal	119607293.92130	66	Sun, 04 Sep 2016 21:43:25
8	↓4	rahul.r	119657214.78911	138	Sun, 04 Sep 2016 18:17:06 (-0.1h)

yPredict (2).csv

SHOW ALL

Results – House Prices Prediction

- Best Model : Ridge Regression
- Pre-processing: None
- Optimal Parameters: alpha = 2.967
- Error on Test Dataset (Mean Absolute Error): 3700



1313	↓150	Shaun D'Souza	0.16876	1	Wed, 05 Oct 2016 07:00:11
1314	↓150	GOT	0.16925	12	Fri, 07 Oct 2016 17:10:33 (-11.4h)
1315	↓150	Sofia	0.16953	1	Wed, 02 Nov 2016 17:28:27
1316	↓150	wizs	0.16975	2	Tue, 06 Sep 2016 01:58:59
1317	↓150	Kaiyue	0.17024	2	Thu, 03 Nov 2016 21:03:51 (-0h)
1318	new	**autoML_iitb**	**0.17046**	**1**	**Sat, 19 Nov 2016 20:19:20**

Your Best Entry ↑

Congratulations on making your first submission!

Tweet this!

1319	↓151	Thalles Domician [BCC]	0.17085	15	Sat, 01 Oct 2016 00:10:15 (-0.1h)
1320	↓151	DongdongWang	0.17087	3	Mon, 07 Nov 2016 03:42:29 (-0.1h)
1321	↓151	Nandita Damaraju	0.17100	1	Tue, 08 Nov 2016 04:40:03
1322	↓151	boudey	0.17102	4	Sat, 01 Oct 2016 14:10:12 (-12.7h)
1323	↓151	Simon Walker	0.17119	1	Tue, 13 Sep 2016 23:14:30
1324	↓151	devin AND j	0.17127	1	Thu, 03 Nov 2016 23:54:53
1325	new	Colin Nattrass	0.17138	2	Wed, 16 Nov 2016 21:58:41

Future Work

- More feature extraction techniques like Kernel PCA, Fast ICA, Polynomial Combination.
- Train a neural network and extract the features from the penultimate layer
- More advanced Regression models and/or Ensemble of models.

THANK YOU!

QUESTIONS?