

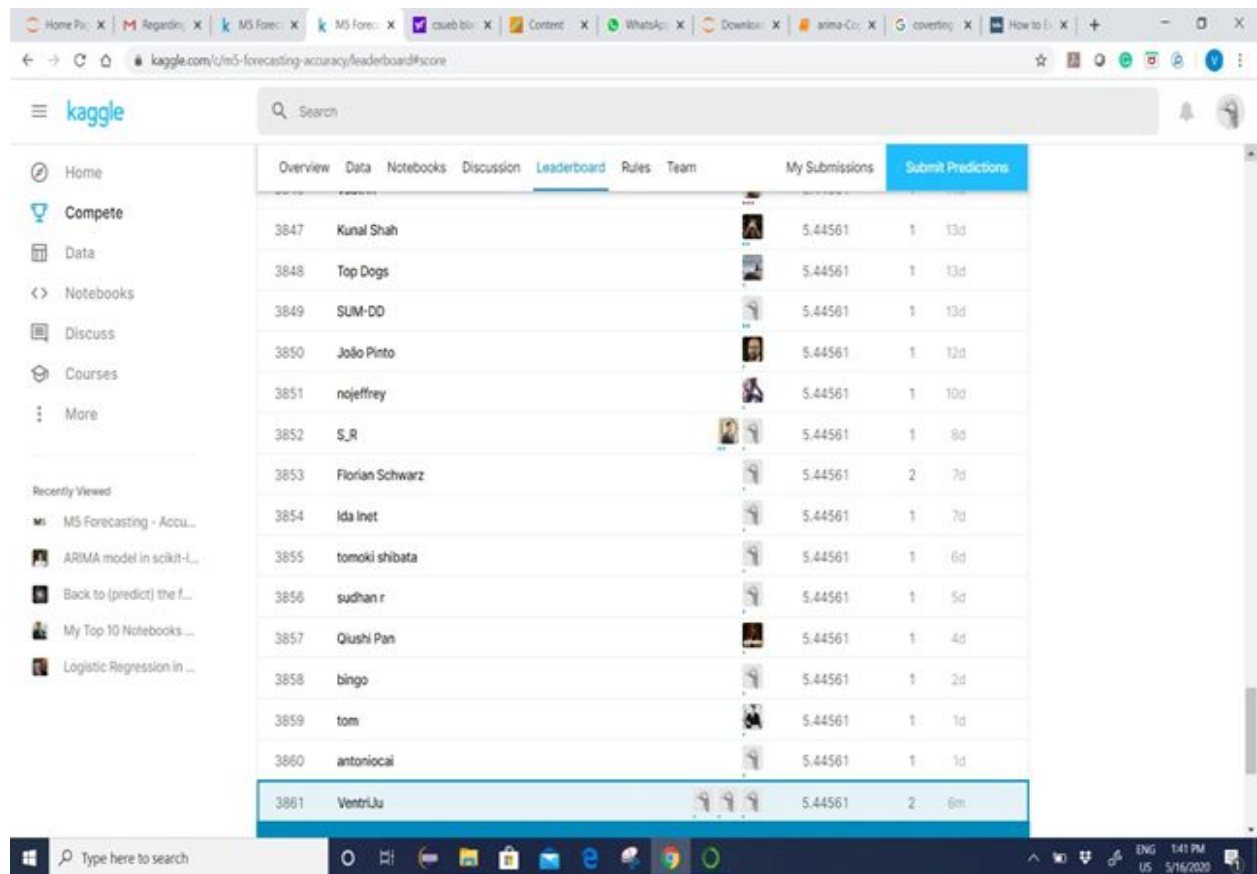
Kaggle Competition “ M5 Forecasting - Accuracy”

Group Name: VentriJu

Members: Vennela Vysyaraju(vq8474), Tripti Prasad(py5697),Juhi Kushwah(tg5354)

Submission1: Score - 5.44561

We used the dataset generated by the Professor for this submission. Have implemented the time series algorithm “Arima” for predicting the sales values for the next 28 days for 1 particular item id ID_HOUSEHOLD_1_001_TX_1. We submitted in Kaggle and received a score of 5.44561. Below is the screenshot of leader board position of our group:



	Overview	Data	Notebooks	Discussion	Leaderboard	Rules	Team	My Submissions	Submit Predictions
3847	Kunal Shah							5.44561	1 13d
3848	Top Dogs							5.44561	1 13d
3849	SUM-DD							5.44561	1 13d
3850	João Pinto							5.44561	1 12d
3851	nojeffrey							5.44561	1 10d
3852	S_R							5.44561	1 8d
3853	Florian Schwarz							5.44561	2 7d
3854	Ida Inet							5.44561	1 7d
3855	tomoki shibata							5.44561	1 6d
3856	sudhan r							5.44561	1 5d
3857	Qlushi Pan							5.44561	1 4d
3858	bingo							5.44561	1 2d
3859	tom							5.44561	1 1d
3860	antoniocal							5.44561	1 1d
3861	VentriJu							5.44561	2 6m

Submission2: Score- 0.62086

We used the original dataset of Walmart for this submission. Have forecasted values for all the items for the next 28 days using Keras categorical embedding V2 and imported tensorflow, regularizers libraries. Have used 3 hidden layers with activation function "tanh". This method improved our score and got a score of 0.62086. Below is the leaderboard position of our group:

The screenshot displays the Kaggle M5 Forecasting Accuracy Leaderboard. The page is titled "kaggle.com/c/m5-forecasting-accuracy/leaderboard#score". The main content area shows a table of participants ranked by their accuracy score. The top participant is VentriJu with a score of 0.62086, followed by Kaz with 0.62005. The page includes navigation tabs for Overview, Data, Notebooks, Discussion, Leaderboard, Rules, Team, My Submissions, and Submit Predictions. A sidebar on the left shows options like Home, Compete, Data, Notebooks, Discuss, Courses, and More. The bottom of the page shows a Windows taskbar with various application icons.

Rank	Participant	Score	Submissions	Time
2120	VentriJu	0.62086	4	22m
2126	Kaz	0.62005	23	12d
2127	yin	0.62030	13	9d
2128	Chris J.T. Auld	0.62038	4	2mo
2129	Mladen Radošević	0.62042	5	1mo
2130	kofuji	0.62060	43	11d
2131	LeonF	0.62080	4	2mo
2132	Akbarali Shaikh	0.62080	5	2mo
2133	Rudy Gilman	0.62082	10	22d
2134	Alberto Gasparin	0.62082	7	2mo
2135	Albedo	0.62082	1	10d
2136	KIM TAE HEON	0.62085	6	5d
2137	[Deleted]	0.62085	1	2mo
2138	Ilija gerasimov	0.62085	1	2mo
2139	blithe_wong	0.62086	1	23d

Submission3: Score-0.62788

We used the original dataset of Walmart for this submission. Have implemented the same as submission2 which is using Keras categorical embedding V2 but with different activation functions like relu, selu and softplus for each and every different hidden layer and also given a different number of neurons for each hidden layer compared to submission2. This implementation gave us a score of 0.62788. This did not improve our leaderboard position and below is the screenshot of our score:

The screenshot shows the Kaggle M5 Forecasting Accuracy Leaderboard. The browser tabs at the top include Documents, 2nd submis..., different ac..., 7 Types of A..., Layer activa..., neural netw..., WhatsApp, csueb black..., Content, and M5 Forecast. The URL bar shows kaggle.com/c/m5-forecasting-accuracy/leaderboard#score. The left sidebar contains navigation links: Home, Compete, Data, Notebooks, Discuss, Courses, and More. Below these are 'Recently Viewed' items: M5 Forecasting - Accu..., Baseline LSTM with Ke..., M5 Forecast: Keras wit..., M5-Forecasting: EDA ..., and M5: Forecasting Tech... The main content area shows the Leaderboard tab selected. The table lists submissions with columns for rank, name, score, and time. Submission 2159 by VentriJu is highlighted in blue, showing a score of 0.62086 and a time of 8m. Below the table, a blue banner states: 'Your Best Entry ↑ Your submission scored 0.62788, which is not an improvement of your best score. Keep trying!'

Rank	Name	Score	Time
2146	yin	0.62030	13 10d
2147	Chris J.T. Auld	0.62038	4 2mo
2148	Mladen Radošević	0.62042	5 1mo
2149	kofuji	0.62060	43 11d
2150	LeonF	0.62080	4 2mo
2151	Akbarali Shaikh	0.62080	5 2mo
2152	Rudy Gilman	0.62082	10 23d
2153	Alberto Gasparin	0.62082	7 2mo
2154	Albedo	0.62082	1 10d
2155	KIM TAE HEON	0.62085	6 6d
2156	[Deleted]	0.62085	1 2mo
2157	Ilija gerasimov	0.62085	1 2mo
2158	blithe_wong	0.62086	1 24d
2159	VentriJu	0.62086	6 8m

Your Best Entry ↑
Your submission scored 0.62788, which is not an improvement of your best score. Keep trying!

Submission4: Score- 0.62086

We used the Naive approach on the original dataset. One thing to observe here is that the score we got for Naive approach is somewhat equal to the score we got while using Keras model with tensorflow. My understanding from this observation is that memory size has a major role to play in analyzing the data in sales_train_validation.csv file. We had to create functions to manage the memory size for our dataset because without this function, the size created hindrance in training and storing values. To understand how this model is forecasting for individual values, we trained 10% of this data, however, instead of doing the cross-fold validation, we kept increasing the values to eventually forecast for 14, followed by 28 days, and while the average value remained unchanged, there was a significant difference in the prediction. This value ranged from [0.2-11.10]. The validation error comes out to be around 0.0883 or 8.83%.

Following is our best score on the leaderboard:-

Overview

Data

Notebooks

Discussion

Leaderboard

Rules


Team

My Submissions

Submit Predictions

2160

Albedo




0.62082

1

11d

2161

KIM TAE HEON




0.62085

6

6d

2162

[Deleted]




0.62085

1

2mo

2163

Ilia gerasimov




0.62085

1

2mo

2164

blithe_wong






0.62086

1

24d

2165

VentriJu



0.62086

8


2h

Your Best Entry ↑

Your submission scored 0.62086, which is not an improvement of your best score. Keep trying!

2166

Tanyapohn



0.62086

15

4d

Submission5: Score- 0.74497

In this approach we used the LSTM model to train past sales values for each 30490 item and a feature which represents if there is an event the following day. Sales of the past 14 days and event feature was used for predicting 15th day sales. Used sales_train_validation.csv file. We are using the last 14 days in order to predict day 1915 sales. In order to predict 1916th day, 13 days from our input data and 1 day from our prediction are used. After that we need to slide the window one by one.

12 days from input data + 2 days from our prediction to predict 1917th day

11 days from input data + 3 days from our prediction to predict 1918th day

....

14 days our prediction to predict the last 1941th day sales.

We submitted and got a score of 0.74497. We did not improve from previous submissions maybe because we used only the past days and one day before event to predict. There are a lot of zero values in the first year. Instead may be multi-step LSTM could have given better results.

Kaggle

Search

Overview Data Notebooks Discussion **Leaderboard** Rules Team My Submissions [Submit Predictions](#)

2155	Albedo		0.62082	1	10d
2156	KIM TAE HEON		0.62085	6	6d
2157	[Deleted]		0.62085	1	2mo
2158	Ilia gerasimov		0.62085	1	2mo
2159	blithe_wong		0.62086	1	24d
2160	VentriJu		0.62086	7	1h

Your Best Entry ↑
Your submission scored 0.74497, which is not an improvement of your best score. Keep trying!

2161	Tanyapohn		0.62086	15	4d
2162	GC Lee		0.62105	57	3d
2163	jrvalentin		0.62113	12	2mo
2164	newman1234		0.62144	6	8d
2165	SaurabhJain		0.62153	6	6d
2166	J_VC200		0.62153	3	2d

Recently Viewed

- LSTM with Keras < 0.7
- M5 Forecasting - Accu...
- Logistic Regression Im...
- M5 Forecast: Keras wit...
- Why time out when cal...

Below is the screenshot of submissions of our group:

Kaggle Competition "M5 Forecasting - Accuracy" | Kaggle

Search

Home Compete Data Notebooks Discuss Courses More

Overview Data Notebooks Discuss Leaderboard Rules Team My Submissions Submit Predictions

13 submissions for VentriJu Sort by Most recent

All Successful Selected

Submission and Description	Public Score	Use for Final Score
submission_wnb.csv 3 hours ago by JUHI KUSHWAH add submission details	0.62086	<input type="checkbox"/>
submission.csv 7 hours ago by tripspras add submission details	0.74497	<input type="checkbox"/>
submission2.csv 7 hours ago by vennela vysyaraju add submission details	0.62788	<input type="checkbox"/>
submission1.csv a day ago by vennela vysyaraju add submission details	0.71208	<input type="checkbox"/>
submission.csv a day ago by vennela vysyaraju add submission details	0.62086	<input type="checkbox"/>
ID_HOUSEHOLD_1_001_TX_1 and TX_2.csv	5.44561	<input type="checkbox"/>

Recently Viewed

- M5 M5 Forecasting - Accu...
- M5 Forecast: Keras wit...
- Baseline LSTM with Ke...
- M5-Forecasting : EDA ...
- M5 : Forecasting Tech...