Predict future sales

Kaggle competition
https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data

Goal

- We are provided with daily historical sales data.
- The task is to forecast the total amount of products sold in every shop for the test set.
- The dataset is the time series dataset. The data consists of 33 months of everyday sales. We need to predict the amounts of items sold in the following (34.) month.

| | shop_id | item_id | | | | | | | | | | | | | | | | | | | item_cnt_day |
date_block_num			0	1	2	3	4	5	6	7	...	24	25	26	27	28	29	30	31	32	33
0	0	30	0.00	3.44	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0	31	0.00	1.57	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0	32	1.50	1.43	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0	33	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0	35	1.00	1.27	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
424119	59	22154	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
424120	59	22155	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
424121	59	22162	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00
424122	59	22164	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
424123	59	22167	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Challenges

- The list of shops and products slightly changes every month.
- Some of items have little amount of occurrences in the sales (e.g. the tweezer was sold in Kogalym shop only once), therefore it will be harder to predict the amount in the following month.
- Original dataset is in Russian language. However, there is additional kaggle challenge with english translation of item names, categories and shops.

Possible approaches

- Linear regression
- Time series forecasting models
    - Autoregressive Moving average
    - Vector Autoregression
- Long short-term memory (LSTM)