



SALES PRICE PREDICTION

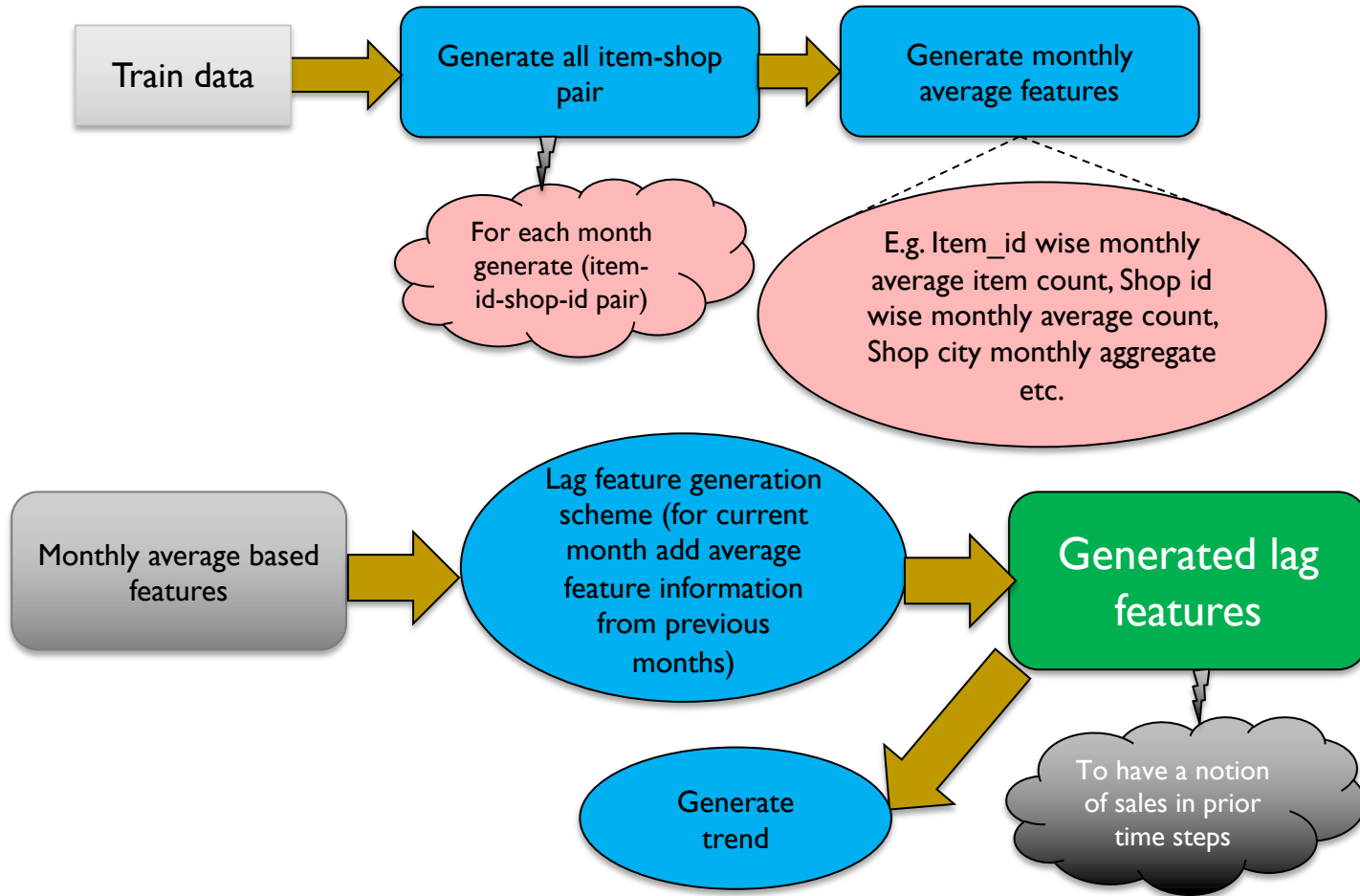
GROUP:CS567_MINREGRESSOR

DIGBALAY BOSE

SOUVIK KUNDU



1. Selecting those train entries with **item_price > 0** (dropping entries with `item_price < 0`)
2. Removing those train entries whose **item_cnt_day** are negative
3. Merging shop entries in training data by shop id whose shop names are same.
4. Splitting the shop name information into **shop category** and **shop city** for each **shop id**.
5. Splitting item category name into two item category based codes (**type** and **subtype**) and combining with **item_category_id** .
6. Splitting item name information into **item name** and **item type**.



What Did Not Work?



MLP

Major steps:

1. Architecture choice
2. Loss function
3. Pre-processing

Possible failure reason:

1. Incorrect normalization
2. Mixture of int and float type features
3. Activation and loss selection



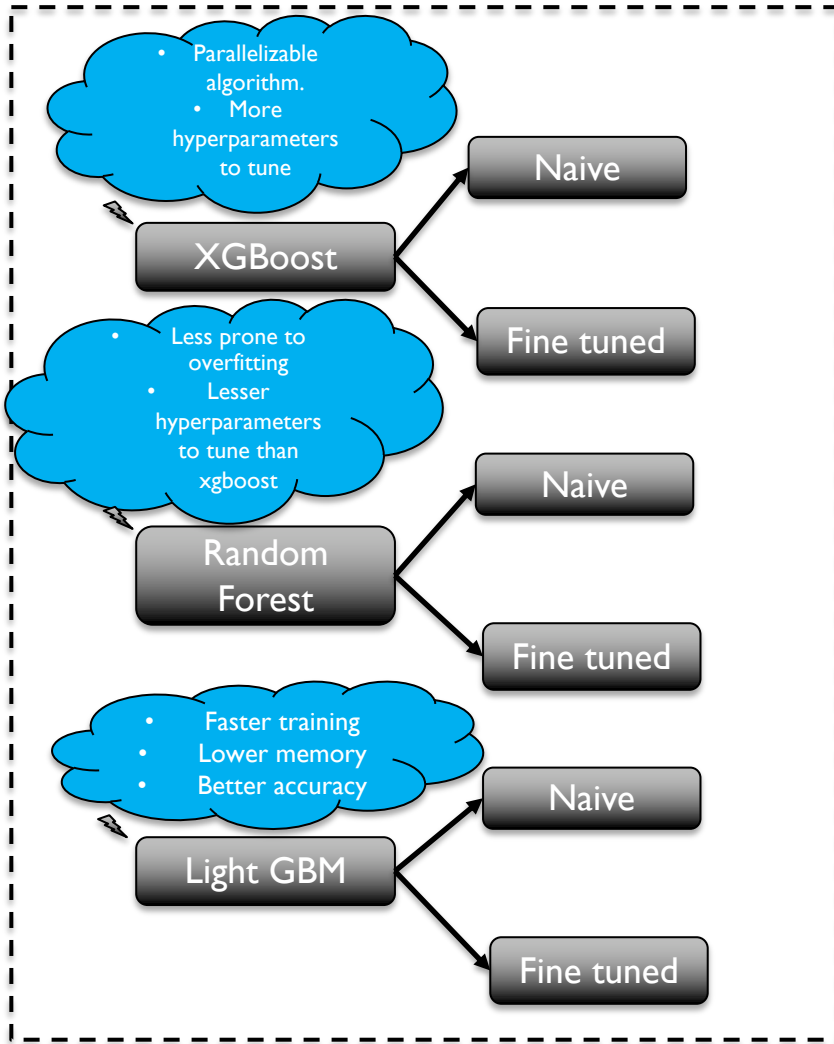
SVR

Major steps:

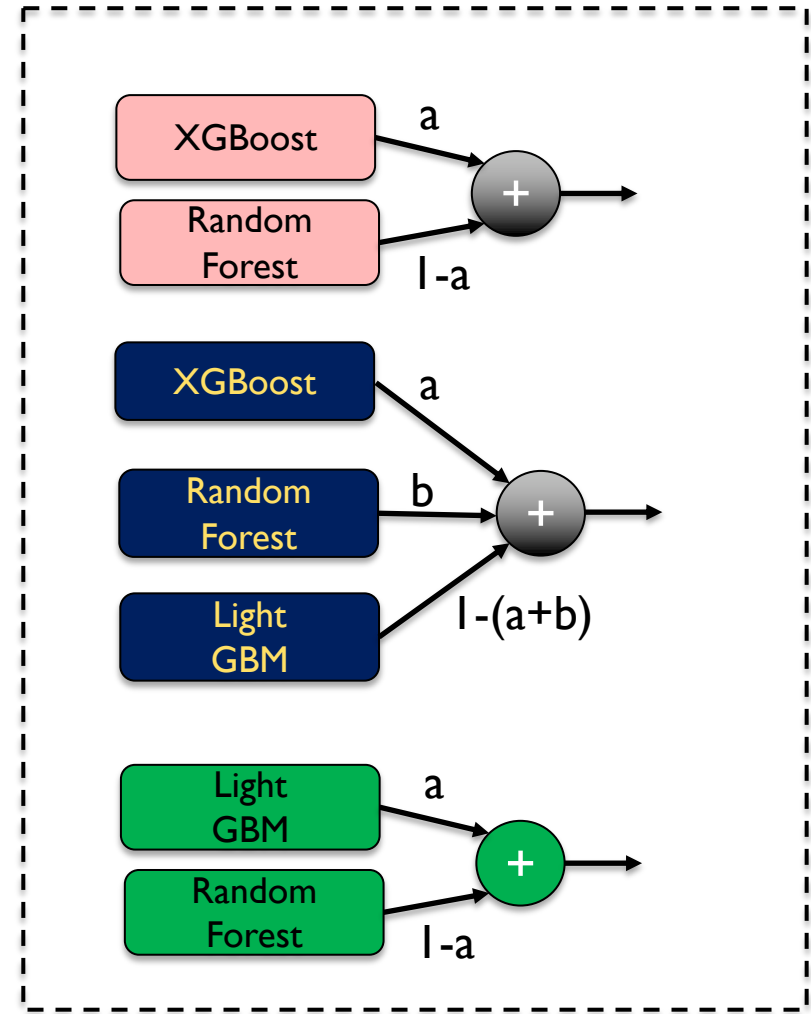
Used scikit learn SVR package

Possible failure reason:

1. Choice of kernel
2. Too time consuming



Stand-alone models



Ensemble models



- Validation data: Entries from month 33
- Test data: Entries from month 34
- Hyperparameter tuning using grid search and hyperopt (<https://github.com/hyperopt/hyperopt>)

Model name	Validation RMSE	Leader-board score
Xgboost		
xgboost-naive(xgb-n)	0.898755	0.9075
xgboost-grid search(xgb-gs)	0.892129	0.89044
xgboost-hyperopt tuning(xgb-hy_tun)	0.891743	-
Random Forest		
Random Forest-naive(rf-n)	0.896694	0.88198
Random Forest-grid search(rf-gs)	0.89587	0.87957
Light GBM		
Light GBM - naive (lgbm-n)	0.904814	-
Light GBM-grid search(lgbm-gs)	0.88636	-
Light GBM (lgbm-hy_tun)	0.8823	-
Weighted Ensemble		
$0.5*(\mathbf{xgb-hy_tun})+0.5*(\mathbf{rf-gs})$	0.887585	0.87808
$0.5*(\mathbf{rf-gs})+0.5*(\mathbf{lgbm-gs})$	0.883526	0.87605
$0.7*(\mathbf{rf-gs})+0.2*(\mathbf{lgbm-gs})+0.1*(\mathbf{xgb-hy_tun})$	0.887478	0.8786

Model details: <https://docs.google.com/spreadsheets/d/1XyucQk70YmVgPNwrti92tOgOz0snbRiawUUWYy0yDvA6U/edit?usp=sharing>



THANK YOU