Applied data scinece

Future sales prediction

1 Introduction

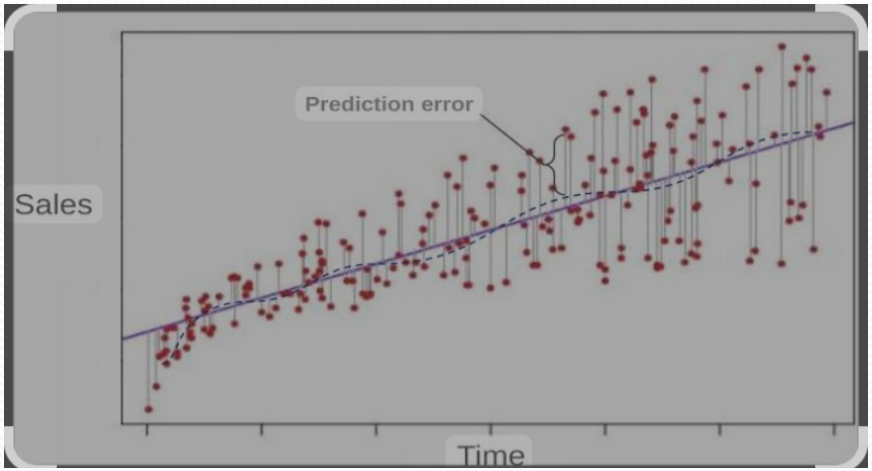
- In this paper we present detailed description and evaluation of the Kaggle future sales prediction challenge
- Training data given in train.csv has 2935849 rows and has 6 columns: date, date_block_num, shop_id,
- The data preprocessing part consists of modifying the outlier elements and merging duplicate information

Feature Engineering

- In order to find the distribution of data across different shops and item categories, we plot the monthly
- In order to extract information from the item category data given in item_categories.csv,
- To extract shop based features, we use the information given in shops.csv. Each entry in shop_name For each month, we first extract the average item count as date_avg_item_cnt.

Problem in future sales predicition

- This series of articles was designed to explain how to use Python in a simplistic way to fuel your company's growth by applying the predictive approach to all your actions. It will be a combination of programming, data analysis, and machine learning



Solution for future sales prediction

- Predictive analytics uses historical data to predict future events. Typically, historical data is used to build a mathematical model that captures important trends. That predictive model is then used on current data to predict what will happen next, or to suggest actions to take for optimal outcomes.
-