

CMPE 257-Team Project Proposal

Project Title : Predicting sales using time series analysis

Team Information : (Team Name: APS)

Team member Name	SJSU ID	Email ID
Akansha Jajodia	014489215	akansha.jajodia@sjsu.edu
Raaga Pranitha Kolla	011824410	raagapranitha.kolla@sjsu.edu
Saranya Visvanathan	014483625	saranya.visvanathan@sjsu.edu

Dataset :

This data is collected from [M5 competition](#) , which is a sales data of 30491X1947 sponsored by Walmart that contains the sales of items over 1941 days. This dataset contains details about sales of multiple items aggregating them into different departments, categories, stores and geographical areas. It is discrete and multivariate time series data. It consists of three csv files which is explained in brief below.

Filename	File Description
calendar.csv	Contains information about the important events.
sell_prices.csv	Contains information about the selling prices for a particular item id with respect to stores
sales_train_evaluation.csv	Contains details about sales for all 1941 days.

Project Motivation & Idea :

Currently, the problems based on the time has been the common one. Time series forecasting provides better performance in many applications like sales prediction, weather prediction, health prediction, etc. It utilizes the past information and helps to provide better predictions on future horizons. Also, understanding and implementing the time series analysis is one of the expected skills for any machine learning engineer. With that motivation, we decided to focus on sales prediction which would happen in the period of 28 days at an e-commerce store (Walmart) based on the historical information of the past sales data. We intend to use time series analysis to find the trends and predict sales.

Project Description :

This project focuses on implementing the following stages of data pipeline such as data preprocessing, exploratory data analysis, data cleaning, feature engineering, modelling, evaluating and comparing ML algorithms and finding predictions about future sales. For data modelling, we plan to use ML learning

algorithms like ARIMA, LSTM, Bidirectional LSTM, CNN, RNN, GRU, LGB models to predict the sales. We plan to implement a minimum of three models (one per person), if time permits, we will work on other algorithms. At the end, all the models would be compared based on the prediction performance.

Tentative project plan :

Task	Responsibility		Execution					
Exploratory Data Analysis	All		We plan to have weekly meetings to discuss our progress and setbacks.					
Data Preprocessing	All							
Modelling	<table><tr><td>Akansha</td><td>LSTM</td></tr><tr><td>Pranitha</td><td>ARIMA</td></tr><tr><td>Saranya</td><td>LGB</td></tr></table>			Akansha	LSTM	Pranitha	ARIMA	Saranya
Akansha	LSTM							
Pranitha	ARIMA							
Saranya	LGB							

For this project, **we require access to high-performance computing system.**

References :

- 1) <https://mofo.unic.ac.cy/m5-competition/>
- 2) Vishwas, B. V., & Patel, A. (2020). *Hands-on time series analysis with python: From basics to bleeding edge techniques*. Berkeley, CA: Apress.
- 3) H. Qu, J. Li and Y. Zhang, "Long Short-term Memory Network Prediction Model Based on Fuzzy Time Series," 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS), Dalian, China, 2020, pp. 417-421, doi: 10.1109/ICAIS49377.2020.9194902.
- 4) Y. Liu, Z. Su, H. Li and Y. Zhang, "An LSTM based classification method for time series trend forecasting," 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 2019, pp. 402-406, doi: 10.1109/ICIEA.2019.8833725.
- 5) Y. Jing, H. Hu, S. Guo, X. Wang and F. Chen, "Short-Term Prediction of Urban Rail Transit Passenger Flow in External Passenger Transport Hub Based on LSTM-LGB-DRS," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2020.3017109.