LITERATURE SURVEY

1.Walmart's Sales Data Analysis- A Big Data Analytics Perspective -
2017 Author-Manpreet Singh, Bhawick Ghutla , Reuben Lilo Jnr , Aesaan F S Mohammed and Mahmood A Rashid

i)Big Data application enables these retail organizations to use prior year's data to better forecast and predict the coming year's sales
ii)Analyzed the data sets of world's largest retailers, Walmart Store to determine the businessdrivers and predict which departments are affected by the different scenarios (such as temperature, fuel price and holidays) and their impact on sales at stores' of different locations.

iii)Made use of Scala and Python API of the Spark framework to gain new insights into the consumer behaviors and comprehend Walmart's marketing efforts and their data-driven strategies through visual representation of the analyzed data

Advantage:

i)Used to understand complex datasets in a matter of time with beautiful visual representations
ii)It help to analyze sale drops which the companies can avoid by using a more focused andefficient tactics to minimize the sale drop and maximize the profit and remain in competition.

Disadavantage:

i)Lack of security since large data processed simultaneously.

2.Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast - 2021 Author - Haichen Jiang; Jiatong Ruan; Jianmin Sun

i)Using time series data to predict future sales changes of products is of great significance to every retailing
company in terms of management and planning of resources.

ii)Analyzes the feasibility of traditional time series model, hybrid models based on time series model and machine learning model, and machine learning model in predicting Walmart sales.
iii)It indicating that the machine learning model performs well in the sales forecast of retail stores. This provides a
new idea for retailers to forecast sales by category and region.
iv)The Prophet model which decomposes trend, season, and holiday and the machine learning model-lightGBM
model- are used to train and test Walmart supermarket sales data

Advantage:

i)The machine learning model prediction method proposed in this paper has smaller prediction errors and results with
good interpretability
ii)seasonal factors and holiday factors can be obtained without feature engineering

Disadvanatge:

i)The traditional time series models cannot predict this data set well
ii)information concerning product, store, and promotion etc. cannot be used when predicting the sales of products

3.Forecasting of sales by using fusion of Machine Learning
techniques - 2017 Author - Mohit Gurnani, Yogesh Korke, Prachi Shah, Sandeep Udmale, Vijay Sambhe; Sunil Bhirud

i)evaluates and compares various machine learning models, namely, ARIMA, Auto Regressive Neural Network(ARNN),
XGBoost, SVM, Hy-brid Models like Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost, Hybrid ARIMA-SVM and STL Decomposition (using ARIMA, Snaive, XGBoost) to forecast sales of a drug store company called Rossmann
ii)models is measured by metrics such as MAE and RMSE. Initially, linear model such as ARIMA has been applied to
forecast sales.
iii)nonlinear models such as Neural Network, XGBoost and SVM were used.
iv)designed using hybrid technique and decomposition technique. Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost,Hybrid ARIMA-SVM were used and all of them performed better than their respective individual models
v)remainder components were forecasted by Snaive, ARIMA and XGBoost

Advantage:

i)Nonlinear models performed better than ARIMA and gave low RMSE.
ii)STL gave better results than individual and hybrid models.
iii)captured linearity and nonlinearity better than ARIMA and ARNN gave the best result of 565 RMSE.

Disadvantage:

i)ARIMA could not capture nonlinear patterns properly
ii)Hybrid Technique can fail if nonlinear model fails to capture residue patterns

4.A Sales Forecast Method for Products with No Historical Data
–2021 Author - Yun Dai, Jinghao Huang

i)In this paper forecasting approach based on the idea of demand for sales forecast of newproducts with no historical data

ii)The results show that compare with a series of machine learning models, the proposed
method increased the forecast performance in a relatively short time

Advantages :

i)By comparing with machine learning models, we find that the proposed model is superiorto others
ii)The proposed method is based on similarity measurement without complextraining so that
forecast can be completed in a short time, and performs well in small-scale data

Disadvantages:

i)the experiment only considers the features of the product and does not consider external influences, such as the impact of regulations on sales.
ii)It uses only small dataset.

5.A Profit Function-Maximizing Inventory Backorder Prediction System
Using Big Data Analytics-2020 PETR HAJEK 1 AND MOHAMMAD ZOYNUL ABEDIN

i)Inventory backorder prediction is widely recognized as an important component of inventory models. However, backorder prediction is traditionally based on stochastic approximation, thus neglecting the substantial amount of useful information hidden in
historical inventory data
ii)It can be achieved by integrating the proposed prot-based measure into the prediction
model and optimizing the decision threshold to identify the optimal backorder strategy.

Advantages:

i)better prediction
ii)profit function performance

Disadvantages:

i)The revenue/loss factors of expected profit are is not used.

6.Planning profitable tours for field sales forces: A unified view on sales analytics and mathematical optimization -2021 Anne Meyer a , Katharina Glock b , ∗, Frank Radaschewski

i)A key task of sales representatives in operational planning is to select the most profitable customers to visit within the next few days .

ii)The final customer selection is strongly interrelated with the tour planning decisions. To this end, variants of the profitable sales representatives tour

problem as a multi-period team orienteering problem .

iii)We show that the selection of a customer scoring and tour planning variant depends on the available data.

Advantages:

i)The search stops when no improvements over the current best solution have been found in 300 iterations

Disadvantage:

ii)Dataset is limited