Machine learning-driven small business Money and inventory management systems

Shardul Dewalkar

14/02/2022

Abstract

A major requirement for small/medium-sized businesses is Inventory Management since a lot of money and skilled labor has to be invested to do so, Also there needs to be a proper machine learning-driven money management system that predicts a fixed amount a businessman can rely on that will be spent or earned in future. There must be a detailed analysis of past business activities such as profits, loss, money invested, expansion percentage. All these things when asked to small/medium businessmen they say they work on an approximation which is where machine learning can be applied using the highly precise model and analysis reports which can improve decision making and make businessmen precisely aware of what has happened in the past with solid figures to compare. This will improve business decision-making and cash flow in more ways than we can imagine! E-commerce giants use Machine Learning models to maintain their inventory based on demand for a particular item.

Inventory Management can be extended as a service to small/medium-sized businesses to improve their sales and predict the demand for various products. When it comes to money management and forecasting, small /medium businesses are completely unaware of any such thing. Demand and income forecasting is a crucial part of all businesses and brings up the following question: how much money was spent this week/month/year? How much income can I expect in the coming weeks/months/years? What was the growth in business this year? What area of business needs improvement? How much stock of an item should a company/business keep to meet the demands, i.e., what should the predicted demand of a product be? Among its many benefits, a predictive forecast is a key enabler for a better customer experience through the reduction of out-of-stock situations, and lower costs due to better-planned inventory and fewer write-off items.

Also, future investment and expenditure with proper income figures can be planned. We discuss the challenges of building an Inventory system and Financial analytics.

1.0 Introduction

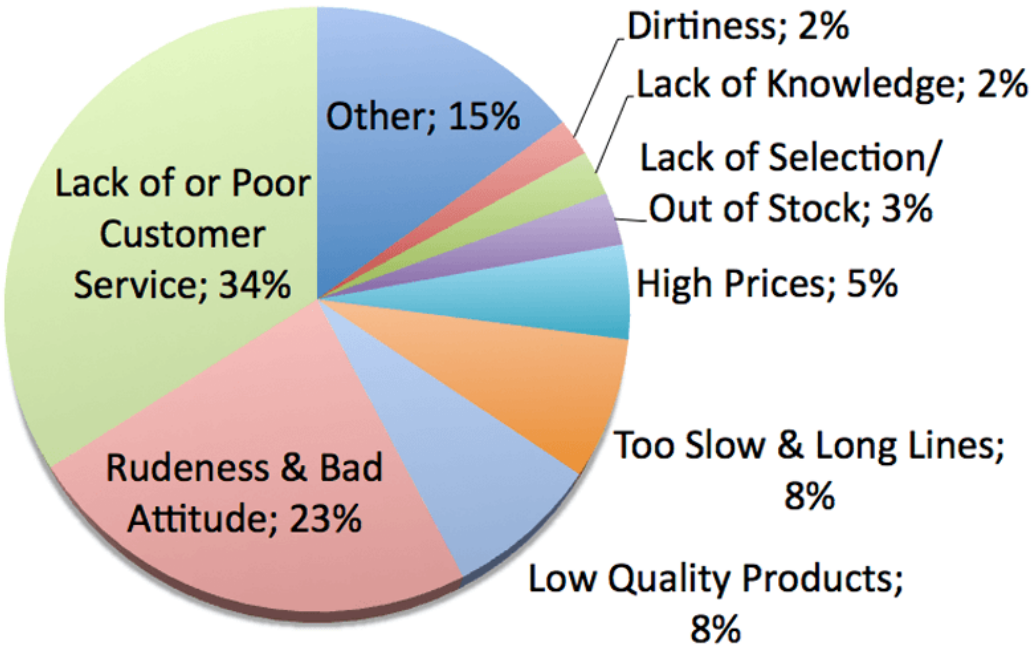
A businessman doesn't know what to expect in the future when it comes to stock that must be available and financial aspects like profit, loss, expenditure, investment, growth. Machine learning when applied in this situation will manage to answer all these questions and also provide valuable insights and correlations which are hard to get otherwise and often get overlooked. All this with a fixed number to compare with past incidents.

Using this model one will precisely be able to make a business decision based on rock-solid evidence and be more reasonable and confident about the approach. he has chosen. He will also be able to compare the impact of his decision on every aspect of business and also with past incidents. Furthermore, he can also take inspiration from decisions taken by E-commerce giants and plan to replicate the same on a small scale to this business.

2.0 Customer Needs assessment:

- Seventy-five percent of millennial shoppers have left a store without purchasing an item and instead bought the item online, according to the 12th annual Global Shopper Survey from Zebra Technologies. Fifty-three percent of Gen X shoppers reported doing the same.
- The leading reason for leaving stores to purchase online was out-of-stock items; 43% of retail associates surveyed cited customer complaints about out-of-stocks as their biggest point of frustration, while 39% of consumers left a store without purchasing due to the issue.

OOS occurs every time an item is not available when a customer would be ready to buy it. Previous studies show that the average Out of Stock rate is about **8%**. That means that **one out of 13 products** is not purchasable at the exact moment the customer wants to get it in the store. This rate rises to 10% or even more about promoted or discounted products.



Small Business TRENDS

Source: GetFiveStars.com
More Charts: http://sbt.me/charts
© 2016, Small Business Trends, LLC

These numbers clearly show that availability of desired product is a big demand of customers and when not fulfilled leads to customer dissatisfaction hence affecting customer loyalty, Increase in competition, business's reputation

Source:

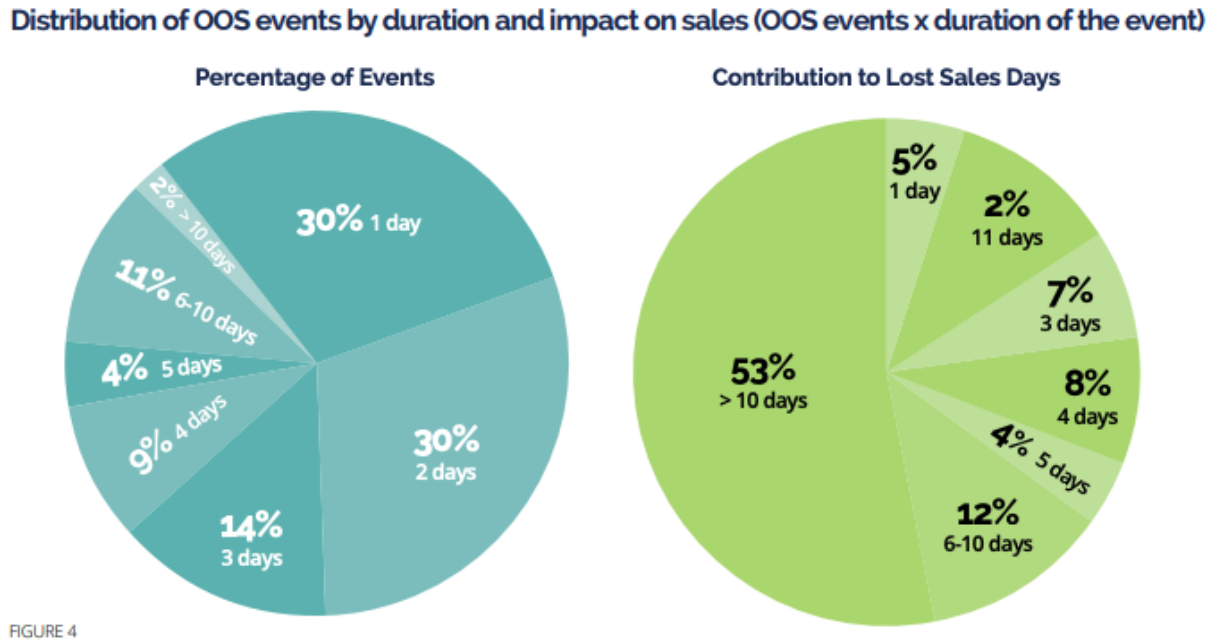


FIGURE 4

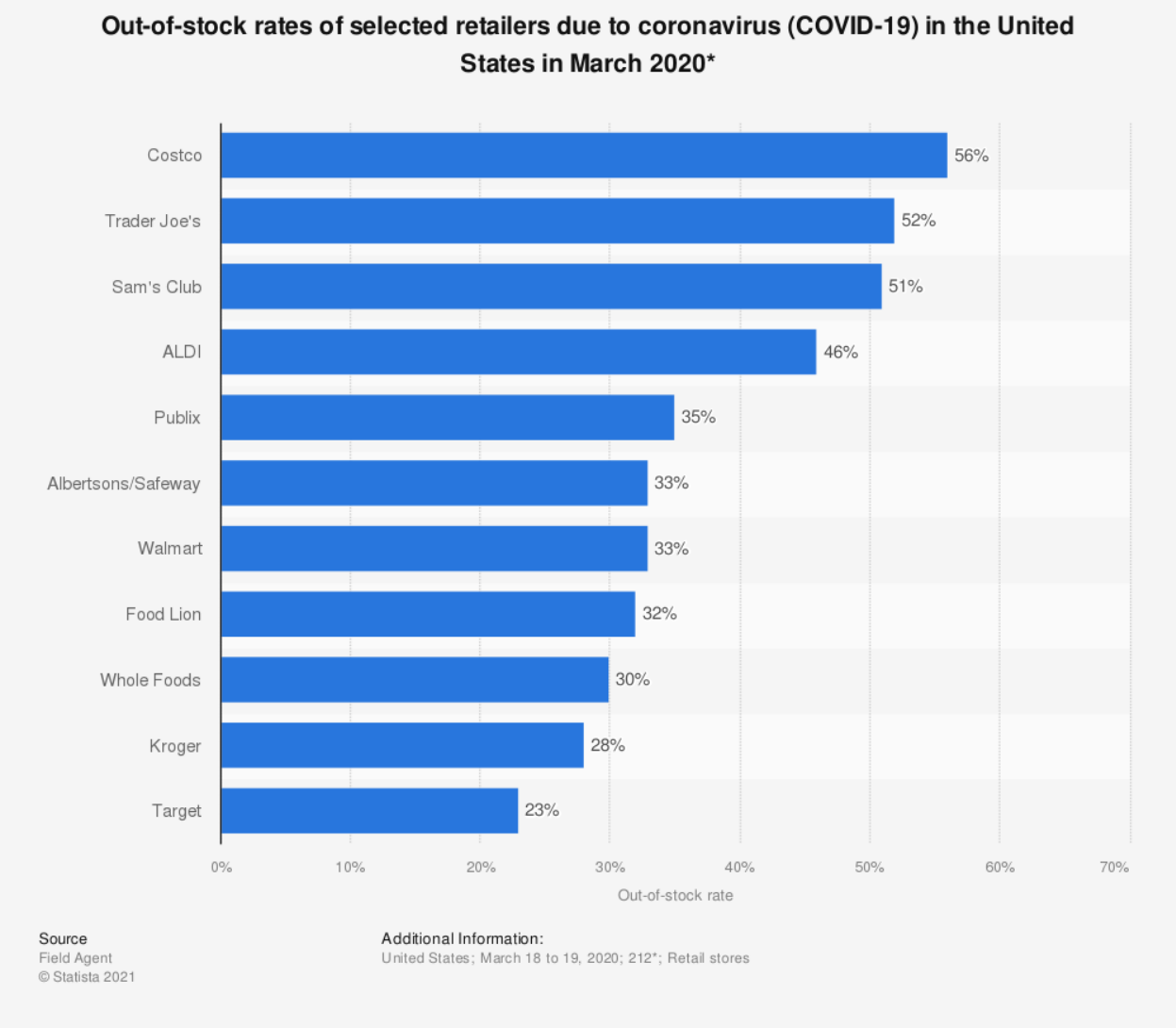
Money management

The survey, carried out by Employers and Business Membership Organizations (EBMOs), involved more than 4,500 enterprises in 45 countries worldwide. EBMOs gathered data from their enterprise members between March and June 2020. The businesses were asked about operational continuity, financial health, and their workforce. GENEVA (ILO News) – A on the impact of the on businesses shows that their greatest challenges have been insufficient cash flow to maintain staff and operations, supplier disruptions, and access to raw materials.

Interrupting cash flow was the greatest problem, the survey found. More than 85 percent reported the pandemic had a high or medium financial impact on their operations. Only a third said they had sufficient funding for recovery. Micro and small enterprises (those with 99 employees or fewer) were worst affected.

At that time, 78 percent of those surveyed reported that they had changed their operations to protect them from COVID-19, but three-quarters were able to continue operating in some form despite measures arising from government restrictions. Eighty-five percent had already implemented measures to protect staff from the virus.

Nearly 80 percent said they planned to retain their staff – larger companies were more likely to say this. However, around a quarter reported that they anticipated losing more than 40 percent of their staff.



With businesses already undergoing significant competitive pressure before the crisis, government restrictions, health challenges, and the economic fall-out brought by COVID-19 further set back many enterprises.

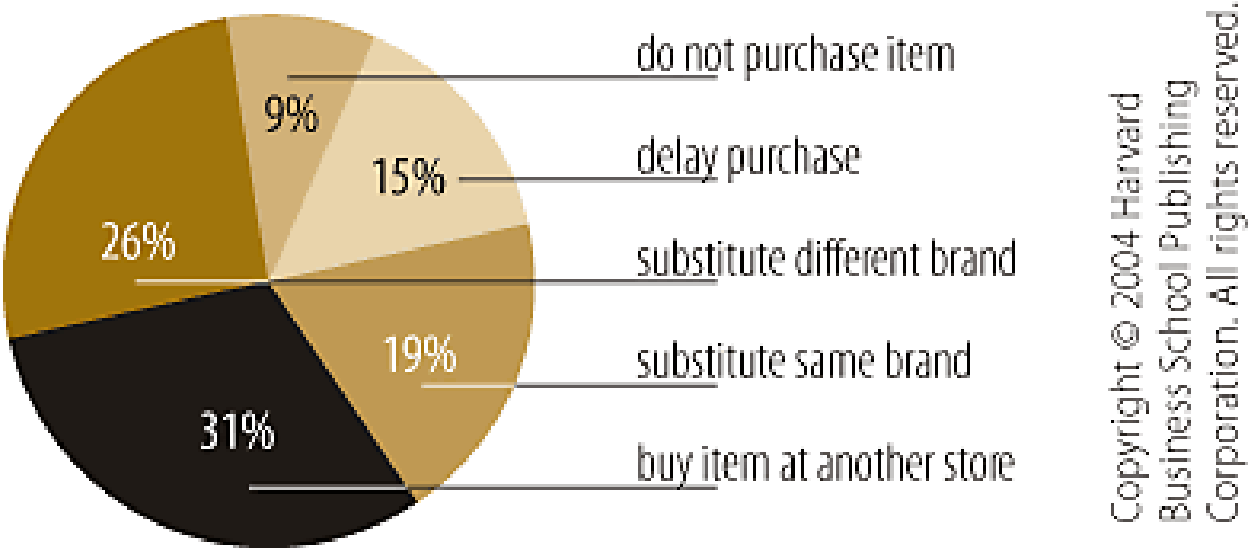
Cash flow was the biggest problem facing businesses during the COVID-19 crisis. Micro and small enterprises worst affected by COVID-19 pandemic, says a survey of 4,500 enterprises in 45 countries.

And that’s how big the management problem is in the business industry. These numbers don’t even account for numerous other small businesses that suffered worst of all in terms of money management. After all healthy cash flow is one of

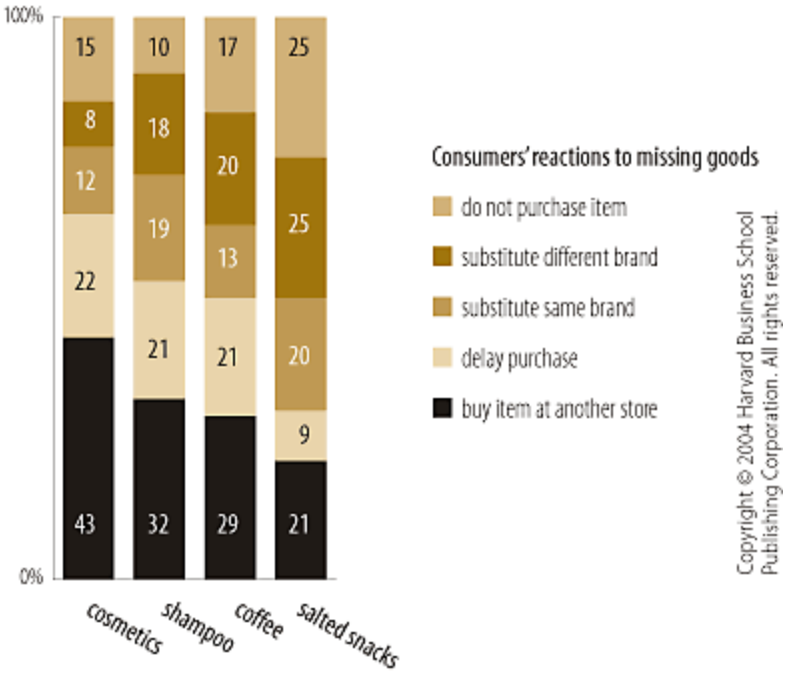
the main factors which keeps a business going. Therefore proper money management systems with strong analytical power are much needed in small/medium industries

3.0 Revised Needs Statement and Target Specification

All these surveys prove that money management and inventory management are needed the most for a small/medium business to grow, make profits, or at least sustain in the market.



Hence the target is to analyze the past data sets make proper reports on what exactly happened in past few years percentage growths, investment returns getting deep into every aspect of business either financial or any other. That will make businessmen see where they are lacking. Proper customer segmentation can be done and accordingly different strategies can be implemented to improve customer loyalty and sales. Several insights can be gained, discover trends and pick out correlations that often get overlooked and make decisions based on all these features henceforth making a businessman more confident in his decision making



because he has concrete figures to support his claims and past data support his future decisions.

4.0 External Search

Sources I have referred for those surveys showing the significance of financial analytics and inventory management on businesses. E-commerce giants have been using inventory management systems but was never been introduced to small/medium scale businessmen. Whereas financial analytics is a completely new concept that is solely targeted towards businesses ranging from small to medium scale. References of all surveys were taken from here:

https://www.netsuite.com/portal/resource/articles/business-strategy/small-business-financial-challenges.shtml

https://smallbusiness.chron.com/overstatement-effects-ending-inventory-21273.html

https://www.thefabricator.com/thefabricator/article/shopmanagement/poor-inventory-management-the-hidden-profitability-killer-in-manufacturing

https://www.retaildive.com/news/survey-39-of-consumers-have-ditched-in-store-purchases-due-to-out-of-stoc/567497/

4.1 Benchmarking

When it comes to money management systems using machine learning there is no such thing yet implemented in the business sector, at least not in small to medium scale businesses. E-commerce giants use Machine Learning models to maintain their inventory based on demand for a particular item but this concept is still not applied to small scale businesses

4.2 Applicable Patents

Since the concept of money management using machine learning is completely new to the market there are no such applicable patents.

when it comes to inventory management there are some patents listed above that can be applicable:

Patent 1-US20050267850A1 - Machine learning systems and methods - Google Patents

Patent 2-US10282722B2 - Machine learning system, method, and program product for point of sale systems - Google Patents

4.3 Applicable Constraints

- Data Collection from business
- Continuous data collection and maintenance
- Lack of technical knowledge for the user(vendors)
- Gaining the trust of businessmen that this product would help boost their profits with detailed analysis report.

4.4 Applicable Regulations

- Govt Regulations for small businesses
- Employment Laws
- Antitrust Regulations
- Regulations against false advertising

4.5 Business Opportunity

Every business needs a proper money management system that can have reliable predictive power gives proper past data analysis reports. Hence this can be applied to all businesses ranging from small to big since this is something new in the market and has great business opportunities. Furthermore, it’s not just limited to business class. any field containing approximation in financial terms can be optimized into whole other level by some modifications in this model.

When it comes to inventory management, Since inventory has only been used by large companies, this can be extended for small businesses, not only shopkeepers or vendors but also food businesses and takeaways. Therefore, there is a fair chance of this service being a great business opportunity. Every small business that depends on sales can and would want to opt for using this service to always know what their customers want. The emergence of every small business is thus a fairly great business opportunity for the service provided by us.

6. Concept selection

The concept of inventory and money management includes a vast set of areas to be dealt with. With evident financial management is the need of the hour.

E-commerce giants define an end-to-end machine learning system using probabilistic demand forecasting models that are built on Apache Spark.

AI proves to be beneficial in handling the customer data and forecasting the purchase behavior of customers which is further extended to sales forecasting which is highly correlated with stock availability. AI can be used to provide notification when a company has to re-order stock and assist in creating a manufacturing schedule considering the variations in demand including seasonal increases accurately and also provide a business with a detailed summary of everything that has happened last week/month containing visualization of every financial aspect of a company whether it be profits, losses, investments, returns. Furthermore, this can also be extended to comparing the current situation with past similar situations or even other businesses going through this. Their good decisions can be replicated in our business. Also, precise forecasting of sales over the next week, month, or even a year is possible using this model when fed with enough data. This will put our business in big advantage over other market competition. The cash flow will be visible and one will also be able to predict what will be this month's income of this business like a corporate job. This will improve decision-making by a great deal.

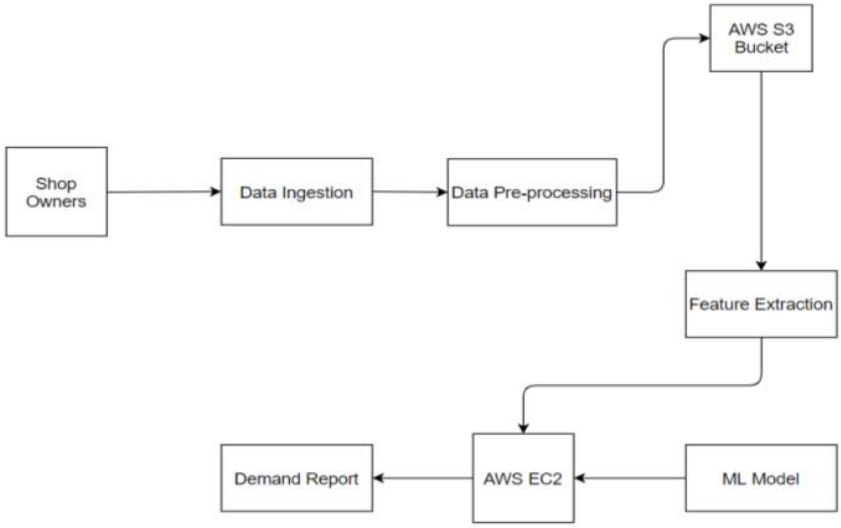
7.0 Final Design

The efficient inventory management system can be achieved by demand forecasting. Demand forecasting is a systematic process of anticipating the demand of a product or service offered by the organization in the future under a set of unpredictable and competitive forces. Whereas sales forecasting and exploratory data analysis is used to understand the current status of business using various

visualization techniques and forecast the profits and loss based on data fed to the model using XGBoost.

In this report, we use the XGBoost regression model to perform demand predictions. But sales forecasting is also possible in a very similar way. XGBoost is a machine learning algorithm that uses decision trees. In prediction problems that have the data, unstructured Neural networks outperform other prediction algorithms, but in our case, the data is structured and tabulated, and decision tree algorithms are considered best for structured data. Hence, we decided to go with the XGBoost algorithm for demand forecasting.

7.1 How does it work?



This whole process of demand forecasting can be further divided into five components: Data Ingestion, Data Pre-Processing, Storage, Feature Extraction, ML model.

A. Data Ingestion:
   Initially, shop owners can log in to the system and add their product details which are stored in a database. Also, the historical sales data is recorded in the database. This data, which is used to train the ML model, is ingested to the system. The training data used consists of several thousands of rows

B. Data Pre-processing:
   Data pre-processing is a data mining technique which is used to transform the raw data in a useful and efficient format. The data ingested is cleaned before it is used for training the model. Hence, the unnecessary fields are eliminated from the database. Pre-processing also involves the transformation of raw data into an understandable format.

C. Storage:
   S3 stands for Simple Storage Service. Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security,and performance. S3 enables customers of all sizes and industries to use it and protect any amount of data for various use cases, such as websites, mobile applications, backup and restore. It is designed for 99.999999999% (11 9's) of durability. The pre-processed data is stored in the S3 Bucket.

D. Feature Extraction:
   The data stored in S3 bucket contains several fields. With the help of proper feature extraction, the accuracy of model can be increased, hence only certain fields of the dataset were extracted for training. For example, the training data consists of product name as well as product ID. Only product ID is required to train the model. Thus, only the required features are extracted from the dataset and used to train the model.

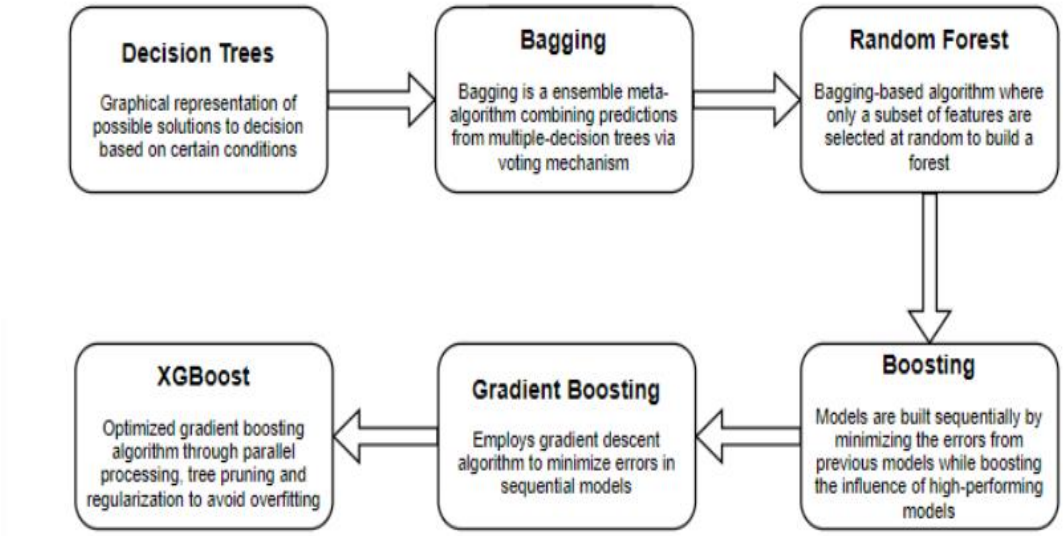
E. ML Model:
   An algorithm called XGBoost is used as the machine learning model. XGBoost is an ensemble algorithm which is based on decision-trees and uses gradient boosting framework. It was evolved from basic bagging algorithm. Normally prediction problems consisting of unstructured data neural networks tend to perform poorly when compared to other

algorithms or frameworks. However, when it comes to small to-medium, structured/tabular data, decision tree-based algorithms perform well.

F. Report:
   The demand values for the next 2 weeks are predicted by the XGBoost model. The predicted demand values are real numbers. To get actual demand values, the values were rounded. The predicted values are output as a report to the shop owners aiding for an efficient inventory management.

IMPLEMENTATION

The model utilized to perform the prediction is known as XGBoost (Extreme Gradient Boosting) regression model. XGBoost is an ensemble decision tree machine learning algorithm. When the data is small and structured, these decision tree-based algorithms outperform other algorithms. XGBoost algorithm was evolved from Bagging Algorithm



The evolution of the XGBoost can be better understood from the figure above, which makes it highly suitable for implementing large amount of data.

A. XGBoost Working:

The XGBoost constructs the decision trees based on the error factor obtained from the previous tree. The technique followed is like that of a normal tree boosting except the way error calculated from the loss function is different. The XGBoost follows certain steps; 1. Model fitting to the data, 2. Model fitting to the residuals, 3. Generating new model (where the new model is boosted version of the old), 4. Repeating all the steps till error is improved.

B. Training the model:

After the model is implemented, training of the model is necessary to predict the demand with better accuracy. The train_test_split is a function for splitting data arrays into two subsets: for training data and for testing data. With this function, eliminates the task of splitting the dataset manually. By default, train_test_split will make random partitions for the two targets x and y. The model is trained using an approach known as early-stopping-rounds which helps to avoid overfitting. It works by halting the training when there

is no increase in the performance on the test dataset up to the specified number of iterations (i.e. up to 100 in our case). It can automatically circumvent overfitting by considering the point of inflection where the performance starts to decrease on test data while for the training data goes on improving towards overfit

Code Implementation

Link:
https://github.com/Shardul1105/ML-AI/blob/8aa397677426af53ade96530ab8bc949824f5daf/Demand%20forecasting.ipynb

In [1]:
```
import pandas as pd
from matplotlib import pyplot as plt
```

In [2]:
```
train = pd.read_csv(r"C:\Users\ashis\train.csv")
print('Shape of the train data:',train.shape)
print(train.head(10))
```

```
Shape of the train data: (913000, 4)
         date  store  item  sales
0  2013-01-01      1     1     13
1  2013-01-02      1     1     11
2  2013-01-03      1     1     14
3  2013-01-04      1     1     13
4  2013-01-05      1     1     10
5  2013-01-06      1     1     12
6  2013-01-07      1     1     10
7  2013-01-08      1     1      9
8  2013-01-09      1     1     12
9  2013-01-10      1     1      9
```

In [3]:
```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 913000 entries, 0 to 912999
Data columns (total 4 columns):
 #   Column  Non-Null Count   Dtype 
---  ------  --------------   ----- 
 0   date    913000 non-null  object
 1   store   913000 non-null  int64 
 2   item    913000 non-null  int64 
 3   sales   913000 non-null  int64 
dtypes: int64(3), object(1)
memory usage: 27.9+ MB
```

In [4]:
```
train.describe()
```

Out[4]:

	store	item	sales
count	913000.000000	913000.000000	913000.000000
mean	5.500000	25.500000	52.250287
std	2.872283	14.430878	28.801144
min	1.000000	1.000000	0.000000
25%	3.000000	13.000000	30.000000
50%	5.500000	25.500000	47.000000
75%	8.000000	38.000000	70.000000
max	10.000000	50.000000	231.000000

In [8]:
```
train.isnull().sum()
```

Out[8]:
```
date     0
store    0
item     0
sales    0
dtype: int64
```

In [9]:
```
test = pd.read_csv(r"C:\Users\ashis\test.csv")
print('Columns of the train data',train.columns)
print('Columns of the test data', test.columns)
```

```
Columns of the train data Index(['date', 'store', 'item', 'sales'], dtype='object')
Columns of the test data Index(['id', 'date', 'store', 'item'], dtype='object')
```

In [10]:
```
test.head()
```

Out[10]:

	id	date	store	item
0	0	2018-01-01	1	1
1	1	2018-01-02	1	1
2	2	2018-01-03	1	1
3	3	2018-01-04	1	1
4	4	2018-01-05	1	1

In [11]:
```
sample_submission = pd.read_csv(r"C:\Users\ashis\sample_submission.csv")
print(sample_submission.head())
```

```
   id  sales
0   0     52
1   1     52
2   2     52
3   3     52
4   4     52
```

In [12]:
```
import xgboost as xgb
# Create DMatrix on train data
dtrain = xgb.DMatrix(data=train[['store', 'item']],
                     label=train['sales'])

# Define xgboost parameters
params = {'objective': 'reg:linear',
          'max_depth': 3 ,
          'silent': 1}

# Train xgboost model
xg_depth_20 = xgb.train(params=params, dtrain=dtrain)
```

```
[20:15:01] WARNING: c:\ci\xgboost-split 1638290375667\work\src\objective\regression obj.cu:188: re
```

```
[20:15:01] WARNING: c:\ci\xgboost-split_1638290375667\work\src\objective\regression_obj.cu:188: reg:linear is now deprecated in
favor of reg:squarederror.
[20:15:01] WARNING: ..\src\learner.cc:576:
Parameters: { "silent" } might not be used.

  This could be a false alarm, with some parameters getting used by language bindings but
  then being mistakenly passed down to XGBoost core, or some parameter actually being used
  but getting flagged wrongly here. Please open an issue if you find any such cases.
```

In [13]:
```
import xgboost as xgb

from sklearn.metrics import mean_squared_error

#dtrain = xgb.DMatrix(data=train[['store', 'item']])
dtest = xgb.DMatrix(data=test[['store', 'item']])


    # Make predictions
train_pred = xg_depth_20.predict(dtrain)
test['sales'] =  xg_depth_20.predict(dtest)

    # Calculate metrics
mse_train = mean_squared_error(train['sales'], train_pred)
mse_test = mean_squared_error(train.sales[0:45000], test['sales'])
print('MSE Train: {:.3f}. MSE Test: {:.3f}'.format(mse_train, mse_test))
```

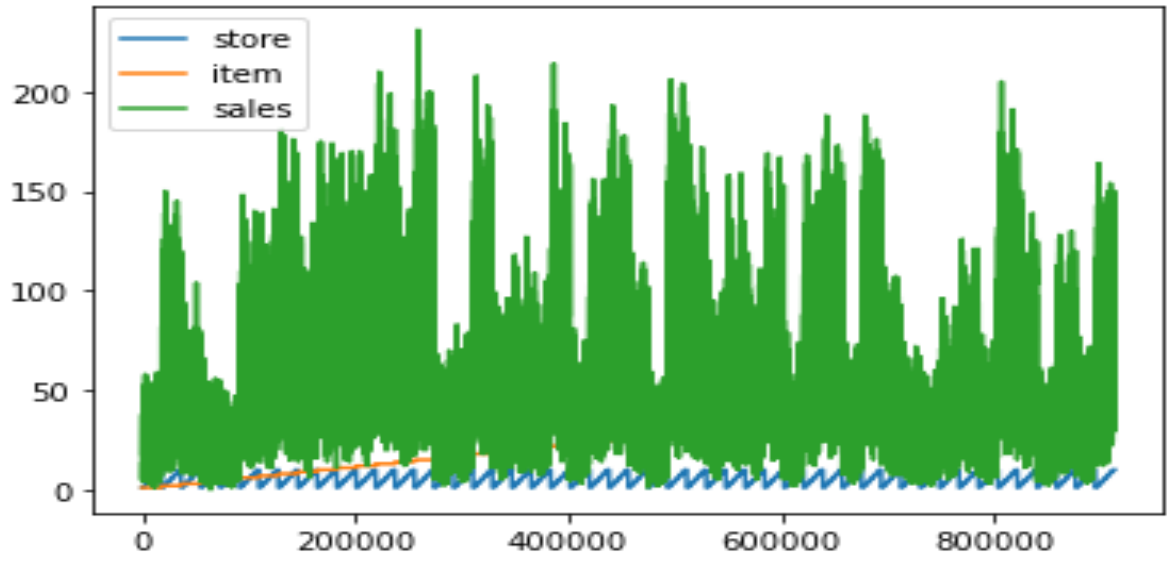
```
MSE Train: 500.043. MSE Test: 815.021
```

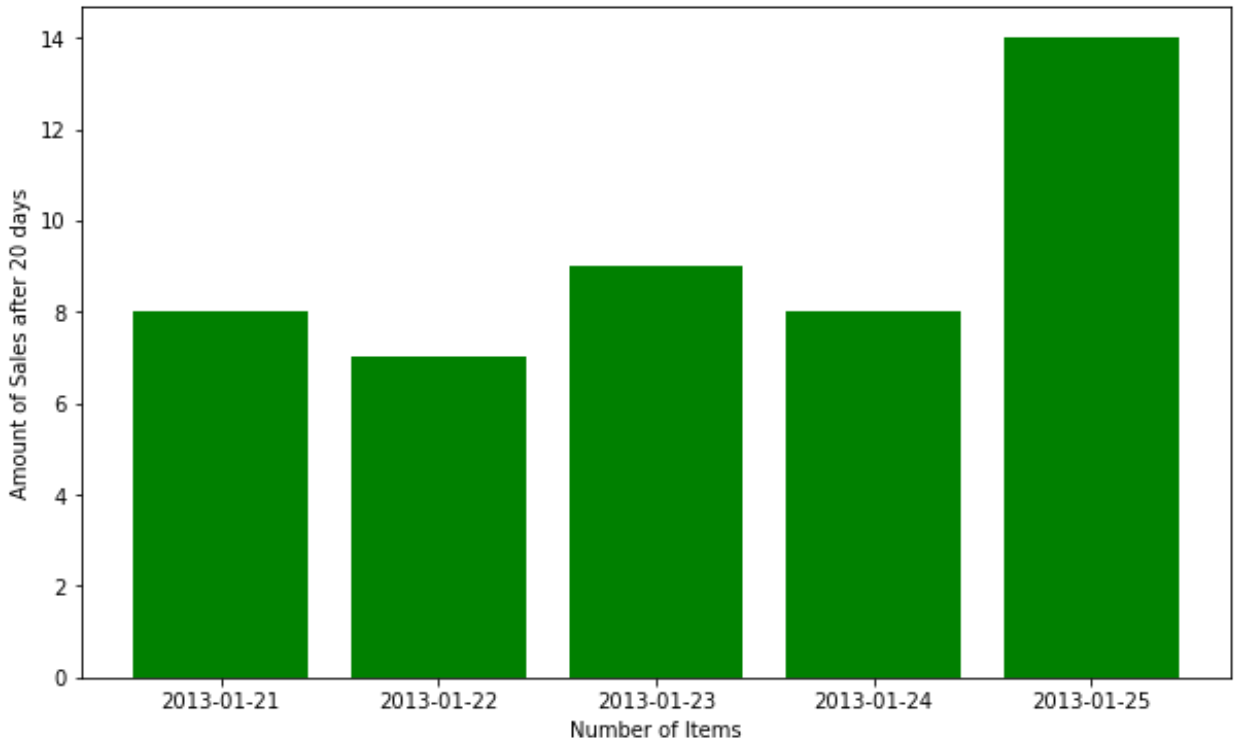
In [14]:
```
print(test['sales'])
```

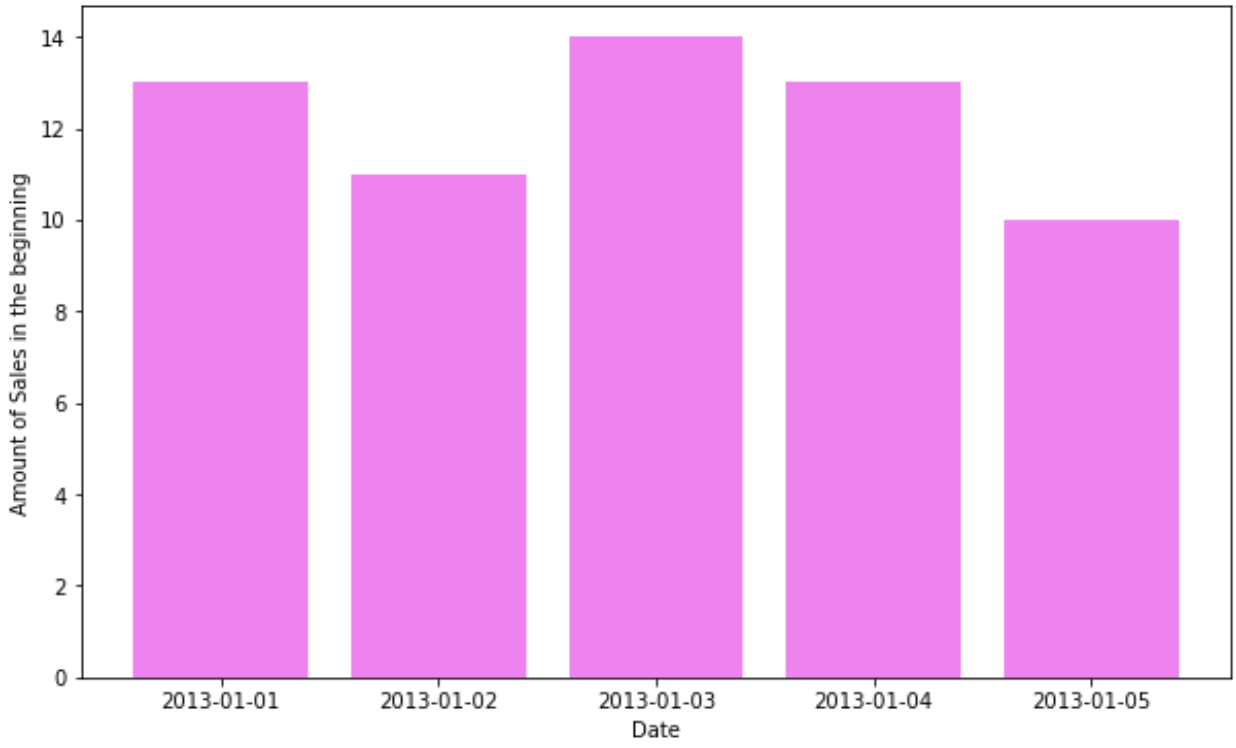
```
0        23.515894
1        23.515894
2        23.515894
3        23.515894
4        23.515894
           ...
44995    60.111858
44996    60.111858
44997    60.111858
44998    60.111858
44999    60.111858
Name: sales, Length: 45000, dtype: float32
```

In [ ]:

Data Visualisation




Amount of Sales after 20 days
0
2
4
6
8
10
12
14
2013-01-21
2013-01-22
2013-01-23
2013-01-24
2013-01-25
Number of Items


Amount of Sales in the beginning
0
2
4
6
8
10
12
14
2013-01-01
2013-01-02
2013-01-03
2013-01-04
2013-01-05
Date

8.0 Conclusions

The Demand Forecasting will help Small/Medium businesses to maintain inventory and minimize manual labour . Allowing it to reduce the capital spent on maintaining inventory. manage business finance at optimum level . In this process it aims at simultaneously improve in profitability . With the forecasting technique, the overstock and stock-out of items are reduced as the stocks are ordered based on the Demand. In the future, the accuracy of the model can be improved by incorporating categorical embeddings in neural networks as it is still a budding topic in the field of neural networks and requires more research.

References

[1]. Joos-Hendrik B¨ose, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger, Yuyang Wang (Amazon), “Probabilistic Demand Forecasting at Scale”, Proceedings of the VLDB Endowment, Volume 10, Issue 12, 2017, pp. 1694-1705.

[2]. Nazar Sohail, Tariq Hussain Sheikh, “A Study of Inventory Management Case Study”, Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 10-Special Issue, 2018, pp. 1176-1190. [3]. Rohaifa Khaldi, Abdellatif El Afia, Raddouane Chiheb, Rdouan Faizi, “Artificial Neural Network Based Approach for Blood Demand Forecasting”, ACM ISBN 978-1-45034852-2/17/03 , 2017.

[3]. https://www.thefabricator.com/thefabricator/article/shopmanagement/poor-inventory-management-the-hidden-profitability-killer-in-manufacturing

[4]. https://smallbusiness.chron.com/overstatement-effects-ending-inventory-21273.html

[5]. https://www.netsuite.com/portal/resource/articles/business-strategy/small-business-financial-challenges.shtml