Project Synopsis

on

**Profit Lifter**

Submitted as a part of course curriculum for

**Bachelor of Technology**

in

**Computer Science**



**Submitted by**

**Gauri Dargar, Aman Malik, Priyambada Pandey**

**1900290120039, 1900290120012, 2000290129004**

**Under the Supervision of**

**Prof. Arti Sharma**

**(Assistant Professor)**

**KIET Group of Institutions, Ghaziabad**

**Department of Computer Science**

**Dr. A.P.J. Abdul Kalam Technical University**

**2021-2022**

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
|  | Page No. |
| TITLE PAGE…………………........................................................................................ | 1 |
| DECLARATION .............................................................................................................. | 2 |
| CERTIFICATE …........................................................................................................... | 3 |
| ACKNOWLEDGEMENT.................................................................................................. | 4 |
| ABSTRACT...................................................................................................................... | 5 |
| Chapter 1. INTRODUCTION |  |
| 1.1.          INTRODUCTION ……………………........................................................ | 6 |
| 1.2 PROBLEM STAEMENT…………………….............................................. | 7 |
| Chapter 2. LITERATURE REVIEW…………………………………………….... | 8 |
| Chapter 3. PROPOSED METHODOLOGY …………………………………........ | 15 |
| 3.1 ALGORITHM PROPOSED | 15 |
| 3.2 FLOWCHART | 18 |
| Chapter 4. OUTCOME OF THE PROJECT…............................................................... | 19 |
| Chapter 5. REFERENCES…............................................................................................ | 20 |

**DECLARATION**

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

**Signature of Students**

**Name:** Gauri Dargar, Aman Malik, Priyambada Pandey

**Roll No.:** 1900290120039, 1900290120012, 1900290129004

**Date:** 12-Dec-2021

**CERTIFICATE**

This is to certify that Project Report entitled “Profit-Lifter” which is submitted by Gauri Dargar, Aman Malik and Priyambada Pandey in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

**Date: 12-12-2021 Supervisor Signature**

**Prof. Arti Sharma**

**(Assistant Professor)**

**ACKNOWLEDGEMENT**

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini Project undertaken during B.Tech. Third Year. We owe a special debt of gratitude to Prof. Arti Sharma, Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for her constant support and guidance throughout the course of our work.

Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. P. K Singh, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

**Signature**

**Date:** 12-Dec-2021

**Name:** Gauri Dargar, Aman Malik, Priyambada Pandey

**Roll No:** 1900290120039, 1900290120012, 1900290129004

**ABSTRACT**

The main aim of this project is to solve problems that small shop owners encounter on the daily basis. One of the major problems that they encounter is the organization of the data of their products and based on that they have to calculate manually which products to buy for reselling to their customers.

this job is very time consuming and takes a lot of effort and despite all the hard work and effort, the retailer can make mistakes which can lead to a financial loss as a small shop owner doesn’t have much stipend to invest, he has to choose the products and their quantity very carefully.

To do this we applied linear regression for the prediction of the items that he should buy which will maximize the profit and save all that hard work and countless hours which he was going to spend predicting manually what and what not he has to buy in a limited capital.

**INTRODUCTION**

Our ML model works on predicting the profit earned by the shop holders on the sale of various items which lead to a computed analysis for the shop holders helping them to purchase those specific items which are profitable to sell and in the other way it reduces the wastage of unconsumed items.

We are using Machine Learning concept it will be implemented in Python Programming language. Implementing supervised learning for the prediction of profitable items.

Small shop owners struggle to decide what products they should purchase for retail and in what quantity,

This is quite hectic and a time-consuming task. On average a small retail shop owner in India spends about 2 hours a day doing the above task and that too is not preciseness. As humans are prone to error sometimes, they make mistakes that can lead them to lose. They may buy something that they are

unable to resell, this is one of the major problems that a retail shop owner face.

As it is quite difficult for a human to analyse thousands of products and make a wise decision. To solve this problem, we applied Machine learning to this task for the shop owner that will recommend products along with their preferred quantities based on the previous data from that shop only. This will maximize the profit of the shop owner and save countless hours of his/her life that he was previously spending in deciding which products to buy and in what quantity.

**PROBLEM STATEMENT**

For the small Low budget Retail Shop holders who cannot afford a computer system to keep track of their sales data. They have to manually keep details accounted in big heavy registers or just keep check of the products in their mind which is quite not a good convenient way.

Sometimes they end up buying products which does not give them valuable profit. Results of such practice leads to wastage of unsold products and the capital the owner invests.

**LITERATURE REVIEW**

**Data science and its relationship to big data and Data-Driven decision making**

The basis for a comprehensive collection of mining strategies data is a small collection of basic ideas that contain scientific data. For data science to grow as a field, instead of being immersed in a river of popular attention, we have to think beyond algorithms, strategies and tools in normal use. We need to think about important goals as well ideas that support strategies, as well as systematic thinking that promote success in data-driven decisions.

These data science theories are common and very broad it works. Succeeding in today's business-focused business environment requires being able to think about how important this is ideas apply to specific business problems – thinking data-analytics. This is helped by the psychologists that they themselves are part of the science of data. For example, the automatic extraction of patterns from data is a process with re-defined categories. Understanding this process and its stages is helpful solving the problem, it makes it more organized, too thus there is very little tendency for error. There is strong evidence that business operations can be greatly improved by data-driven decisions, 3 large data technology, 4 and data science strategies are based on greatness data.9,10 Data Science supports data-driven decision-making and sometimes allows automatic decision-making to a large extent - and depending on the technology of the ‘big’ data’’ storage and engineering. However, the terms of scientific data is theirs and should be viewed and discussed openly so that the science of data can see it power.

**How is Statistical Engineering Different from Data Science?**

Developing evidence-based approaches to solving complex problems is the key a major focus in the field of mathematical engineering, and in our view it is very important to add to both applied statistics and data science. These methods can serve as an umbrella for to solve problems related to statistical (descriptive) status data and data natural (predictable) science for large and small data sets.

Improving continuity Education courses for existing data professionals in mathematics, engineering, mathematics, and data science is also an important step for ISEA. In addition, ISEA is partnering with universities to develop courses to teach the next generation of problem solvers.

**Machine Learning Types**

Machine-learning has evolved and recently there is a tremendous increase in its applications in the department of medical science .one such department is radiation oncology. in this paper, we focus on ML for radiation outcome modeling. it includes survival analysis, local tumor control probability (TCP), and normal tissue control probability (NTCP) In the past, we only had analytical models in which we suffered when it is no consistency in the underlying assumptions and simplifications. while ML algorithms can deal with these problems without suffering and predict accordingly

MACHINE LEARNING APPROACHES FOR RADIATION OUTCOME MODELLING

FOR STRUCTURED DATA for structured data the algorithms like linear regression, artificial neural network, support vector machines, BNs, DTs, RFs, or gradient boosting machine (GBM). now an important question arises which algorithm or algorithms should be preferred. we know that no single algorithm is best for all problems. however, generally, we can show that a particular algorithm will be best for a specific problem. for example, if there is enough and we need accurate results the RF and GBM are on average best algorithms for the analysis of structured data. If our goal is interpretation then DTs, BNs, linear models can be used.in the case of radiation oncology the datasets are smaller than those in other field sit was proved that on average linear models, RF, GBM provides the best results therefore it is advised that at least these algorithms should be explored in the case of structure data.

FOR UNSTRUCTURED DATA in many parts of the medical datasets in radiation oncology, the data is in unstructured format. to deal with this kind of data deep learning can be used. it is a subdomain of machine learning. CNN's can be used to predict the survival risks for cancer patients based on imaging data. Also, DNNs can be used to calculate state transition probabilities for building an autonomous clinical decision support system to adapt patient dose per fraction in a response-adapted treatment setting. deep learning methods in general are more promising than any other algorithms in the case of unstructured data. but one problem with these algorithms is that they need a fairly large dataset to provide promising results. therefore, it is advised to proceed with caution in case of small datasets.

**A Few Things to Know About Machine Learning**

The Article on few things to know about Machine Learning gives us the folk knowledge which is must needed to advance machine learning applications. As known to us earlier machine learning system automatically learns programs from the data provided by the user. In this era machine learning is used everywhere be it spam filter, drug design, web search, fake news detection and many other applications. Machine learning which is also known as Data Mining or predictive analytics figures out how to perform important tasks by generalizing from examples and as more data is available more ambitious problems can be tackled. The article summarizes 12 key lessons which must be kept in mind for a machine learning researcher or practitioner.

The learning phase of a machine model includes three things which are representation, evaluation and optimization. There are bewildering variety of learning algorithms are available, a classifier must be represented in some formal language that the computer can handle. Choosing the set of classifiers that it can possibly learn this set is called Hypothesis space of the learner. The evaluation function helps to distinguish good classifier from the bad ones. Finally, the need of a method to search among the classifiers in the language for the highest scoring one this technique is known as optimization. The fundamental goal of machine learning is to generalize beyond the training set. As in the earlier years machine learners split the dataset into training and test data which helps to generalize beyond the training set but with more advancement other methods like cross validation are used for more successful results.

**Transformer Neural Networks (Attention is all you need)**

The best performing models also connect the encoder and decoder through an attention mechanism.  
Attention Is All You Need is sequence of recurrent or convolutional neural networks that include encoder and decoder so far, the best performing model.

Recurrent neural networks architectures:

1. Vector – Sequence Models: These NNs take in a fix size vector as input and give output sequence of any length, in image Captioning the input can be a Vector of image and output can be a sentence describing the image.
2. Sequence – Vector Models: These take in a sequence as input and spits out a fix length vector, in sentimental analysis takes in a sequence of characters as input giving fix vector as result.
3. Sequence -Sequence Models: It takes a sequence of input and gives out a sequence of output used in language translator for example it takes a sentence in French as input and translates in English.

Disadvantages with RNN:

Slow to train so we use truncated backprop through time.

Long sequences lead to vanishing/exploding gradients when network is too long.

LSTM Networks:

It is long short-term memory networks it has a memory cell which allows past information to skip processing of the past neurons so memory is retained for longer sequences as the LSTM models are more slow than normal RNNs they take input in series format i.e., one word after the other.

To parallelize the sequential data: Transformers were introduced in 2017, similar to RNNs but the difference is that the sequence can be passed in parallel. Consider example to convert French sentence to English in RNN we consider one word after another, the current word hidden sate has dependences in the previous words and the word embeddings are generated one at a time where as in Transformer encoder

 How it works?

Brief idea of Transformer components.

Input Embedding: As the computer does not understand words, we use the concept of input embedding where we map every word to a point in space where similar words are grouped together this space is called Embedding Space this space marks a word to a vector but same word in different sentence may have different meaning, this is where positional encoders come in the picture. Positional Encoder: It is a vector that gives context based on position of word in sentence. Previously used a sin and cosine function to generate this vector. So after applying embedding and then adding the positional encoding we get word vector which have positional information.

Then we pass this to the encoder block where we get a Multi-Head Attention layer and a Feed Forward Layer. Attention: It focuses on what part of the input should we focus at a time. if we are translating from French to English and doing Self Attention that means attention wrf to oneself. We focus on how relevant ith word in the English sentence does is to other words in the sentence which is represented in the Attention vector. Feed Forward Network: This is a simple layers of feed forward network applied to every one of the attention vectors they transform the attention vectors into a form that is digested by next encoder/decoder block. Decoder: It also follows the basic steps of converting the French language to into positional vector which is then passed to decoder block where there are Three components

The self-Attention Block generates attention vector these vectors and attention vectors from encoder block are then passed into another Encoder-Decoder Attention which will see how related each sentence in French and English are. This is the main mapping takes place, then it is passed to feed forward layer to make it more digestive. then Linear Layer which is again a feed forward layer used expand dimensions and the SoftMax transforms into Probability distribution and the final word is the word correspond to the word with the highest probability.

**A Unified Approach to Interpreting Model Predictions**

There are many recent efforts in providing interpretability to complex models, such as LIME, DeepLift and Layer-wise relevance propagation. In complex models, change in input features contributes to the output change nonlinearly. Many previous works tried to distribute the feature contributions fairly. In this work, the author pointed that Shapley value -- marginal contribution of each feature -- from game theory is the unique solution that satisfies desired properties. The author provides a unified framework for many of the interpretable methods by introducing the concept of additive feature attribution methods and SHAP as a measurement for feature importance. A SHAP value estimation method is provided. The author supported the method with user studies and comparisons to related works on benchmark datasets.

This paper is technically sound with detailed proofs and supportive experiments. I think there could be more discussion on the shapely value estimation accuracy vs computational efficiency.

This work provides a theoretical framework for feature importance methods. The

better defined metric SHAP could become a guidance of feature importance evaluation in the research field of interpretable methods.

**Probabilistic forecasting of heterogeneous consumer transaction-sales time series**

Key desiderata are to define full probabilistic forecast distributions for each of many items at the level of individual stores and departments within stores, with a focus on daily sales forecasting over multiple days ahead at each time point. The aim is to do this with a model class that is flexible enough to be tailored to individual products, so as to address the enormous diversity experienced in daily sales across many thousands of supermarket items over large numbers of stores in supermarket chains. Such models must integrate and account for various levels of seasonality, item-level covariates , and otherwise allow for and adapt to unpredictable drifts in levels and variability of sales as they arise. Challenges in daily sales forecasting at the store level begin with many items that sell sporadically,   
 the so-called intermittent demand problem generating many days with zero sales for such items. A full probabilistic model must define time-adaptive, item-specific probabilities of zero/non-zero sales patterns, and forecast accuracy assessment must include relevant metrics for probabilistic predictions. A second challenge is that of potential high variability and extreme values in daily sales of items that do sell more frequently, features that have been addressed using various modified Poisson, negative binomial, jump-process models, and others.

The modeling advances in this work capitalize on availability of detailed point of sale data on transactions and sales-per-transaction information on supermarket items. Data are observed daily with day t records of the number of transactions involving this item, i. Many other items sell more frequently but again generally at 1 or perhaps 2 units per transaction. Then other items can sell at higher levels per transaction, though again generally small numbers.   
  
Infrequent bursts of item sales occur, often in the context of known promotions or pricing changes. Some items experience rare events in terms of larger numbers of sales in rare batch purchases. Bt is the number of transactions– or baskets– involving at least one unit sale

First, we utilize a dynamic count mixture model to represent and forecast the item-specific transaction process bt over time. This class of DCMMs provide a flexible framework for modeling non-negative counts that is customized to dealing with zero counts together with potentially diverse patterns of variation of non-zero counts. Each model component may involve covariates– such as price and promotion predictors, seasonal effect variables, holiday effects, and so forth– that may partly explain and hence predict variation over time in transaction outcomes. An initiating application for the development of DCMMs was in forecasting item sales, and one important aspect of these models is that they naturally integrate time-specific random effects. The key point here is to adapt DCMMs to model transactions, not sales. The heterogeneity and overdispersion seen in sales data is, in part, due to the compounding effect of varying size of transactions per customer throughout the day.   
  
Some aspects of variation over time– in both zero/non-zero transaction probabilities and in the conditional levels of non-zero transactions– comes through the specification of covariates in the regression vectors. Additional aspects of variation can be captured

adjusted for through time   
  
Many items sell just once per transaction, many others sell at perhaps 2 or 3 items, with higher numbers becoming increasingly rare. The multi-scale formulation of a DBCM is motivated by the reality that predicting rare events of any kind– here, larger numbers of units per transaction– is only and properly addressed using hierarchical sequences of conditional probabilities to define chances of outcomes. The DCMM defines forecast distributions for transactions bt into the future, and is used to compute predictive probabilities of transaction outcomes as well as– critically– to simulate representative future outcomes. Monte Carlo analysis, repeatedly simulating many representative values of bit and then sales coupled to each value defines formal computation from the required

redditive distribution of sales. As we move across Monte Carlo samples, uncertainty about transaction levels is represented, and then the conditional uncertainty about sales per transaction factors in.  
  
In most applications, it is of interest to use direct/forward simulation of multi-step ahead predictive distributions. Among other things, this allows trivial computation of probabilistic forecast summaries for arbitrary functions of the future data over multiple steps ahead. In transactions and sale forecasting, generating Monte Carlo samples of synthetic futures over a series of days provides forecast summaries for sales each day, the patterns of variation and dependence day-to-day, and other aspects of applied relevance such as cumulative forecasts over a period of days. , the generation of multiple random samples of transactions and sales outcomes over multiple days, defining «synthetic» futures that can be summarized to compute a range of point forecasts of interest under various utility functions, as well as full probabilistic summaries that formally capture and reflect predictive uncertainties.

Each row in the transaction-level data set represents one consumer’s purchase of one or more units of a single item. Items are identified by a unique base universal product code in the «Dry Noodles and Pasta» category. For each transaction event, the data includes item UPC, the purchase date, the effective price per unit, whether or not the item was purchased on promotion, and the unit sales in the given transaction. These items represent a range of transactions-sales patterns and typify the features of data across many items.   
  
Both series also share the feature of somewhat rare extreme values, although the diminished variability of the transaction data is evident.

**Predicting Online Product Sales using Machine Learning**

Product sales prediction is a major aspect of purchasing management. One of the key challenges faced nowadays by organizations the dynamic, international and unpredictable business environment in which they operate. Sometimes decision regarding whether or not to make a purchase is dependent on price but in many cases the purchasing decision is more complex. Retailers nowadays understand this well and attempt to make use of it in an effort to gain an edge in a highly competitive market.

Dataset which was used in the project was 2-3 years old and the sales which it was predicting now was on the basis of that data. Nowadays data is being generate data such a large rate and there would be so much changes in new data in comparison with the stored data

Forecast for big market sales This approach was proposed by Deven Ketkar. In this methodology raw data collected at big mart was preprocessed for missing anomalies and outliers. Then an algorithm was trained on this data to create a model Algorithms used were Random forests and multiple Linear Regression

SALES TIME SERIES FORECASTING This approach was proposed by Bohdan M. Pavlyshenko. This methodology is a stacking approach for building regression. Ensemble of single models was studied for implementation.

The main objective of the project is to show that product demands can be predicted through the comparative influence of promotional marketing strategies such as discounts and the provision of free delivery choices, user generated contents such as volume and valence of on-line reviews, and sentiments of the web reviews

The paper proposes that promotional marketing strategies and social interactions such as online review and answered questions are both important for influencing sales. In summary, we have shown that when sentiments interact with volume and valence, it becomes a more important predictor of product sales.

**PROPOSED METHODOLOGY**

**Algorithm Used:**

Among the various machine learning algorithms used in supervised learning which are

* Linear Regression or multivariate regression
* Logistic Regression
* Nearest Neighbor
* Gaussian Naive Bayes
* Decision Trees
* Support Vector Machine (SVM)
* Random Forest

We will be using our dataset to work on most of these above Proposed algorithms to compute the best results among them

**Most considered is Linear and multivariate regression model**

**Assumption of Linear regression**

* Linear relationship between dependent and independent variable
* All variables of regression to be multivariate normal
* Particularly there is no or little multicollinearity in the data
* Response variable is continuous and also residuals are almost same throughout the regression line

For error calculation the method of Least square (a standard approach) is used to determine the best fit line for the given data

coefficients a and b

 y hat is Predicted value



yi is actual value

Multiple Linear Regression

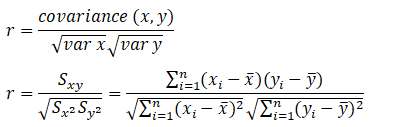
It is a statistical technique that uses various explanatory variables to predict the outcome of response variable.

Formula is y=b0+b1\*x1+b2\*x2+.......bn\*xn Where y=dependent variable and x=independent variables

## Correlation Coefficient

The linear correlation coefficient r measures the strength of the linear relationship between the paired x and y values in a sample.

Pearson’s [Correlation Coefficient](https://sixsigmastudyguide.com/linear-regression/)



Note that -1≤ r ≤ +1

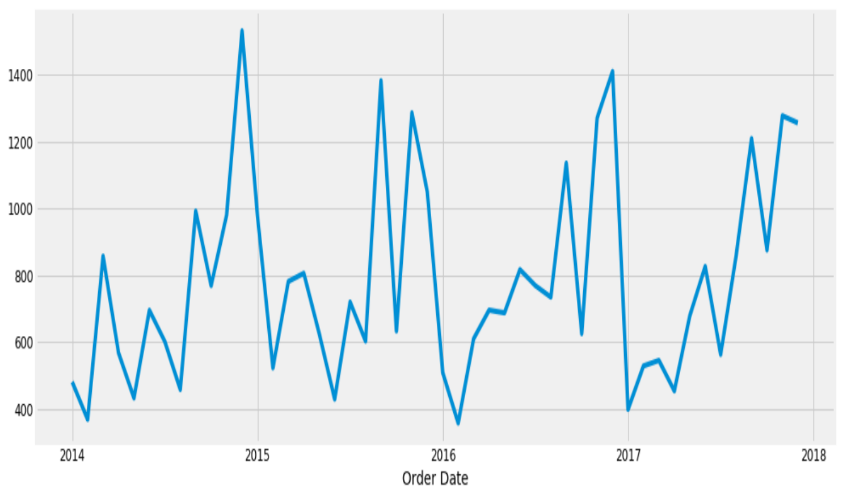
* The line slopes upward to the right when r indicates positive value
* The line slopes downward to the right when r indicates negative value
* A value closer to 1, indicates the stronger positive linear relationship
* A value closer to -1, indicates the stronger negative linear relationship
* When r=0 implies no linear correlation

**Dynamic Data**

If our dataset comprises of time series data which has a variable mean depending upon time period.

Time series are widely used for non-stationary data, like economic, weather, stock price, and retail sales in this post. We will demonstrate different approaches for forecasting retail sales time series.

This is what time series data looks like



Similarly, the data for sales of retail items can be a time series data with variation in sales of certain products The time-series has seasonality pattern, such as sales are always low at the beginning of the year and high at the end of the year.

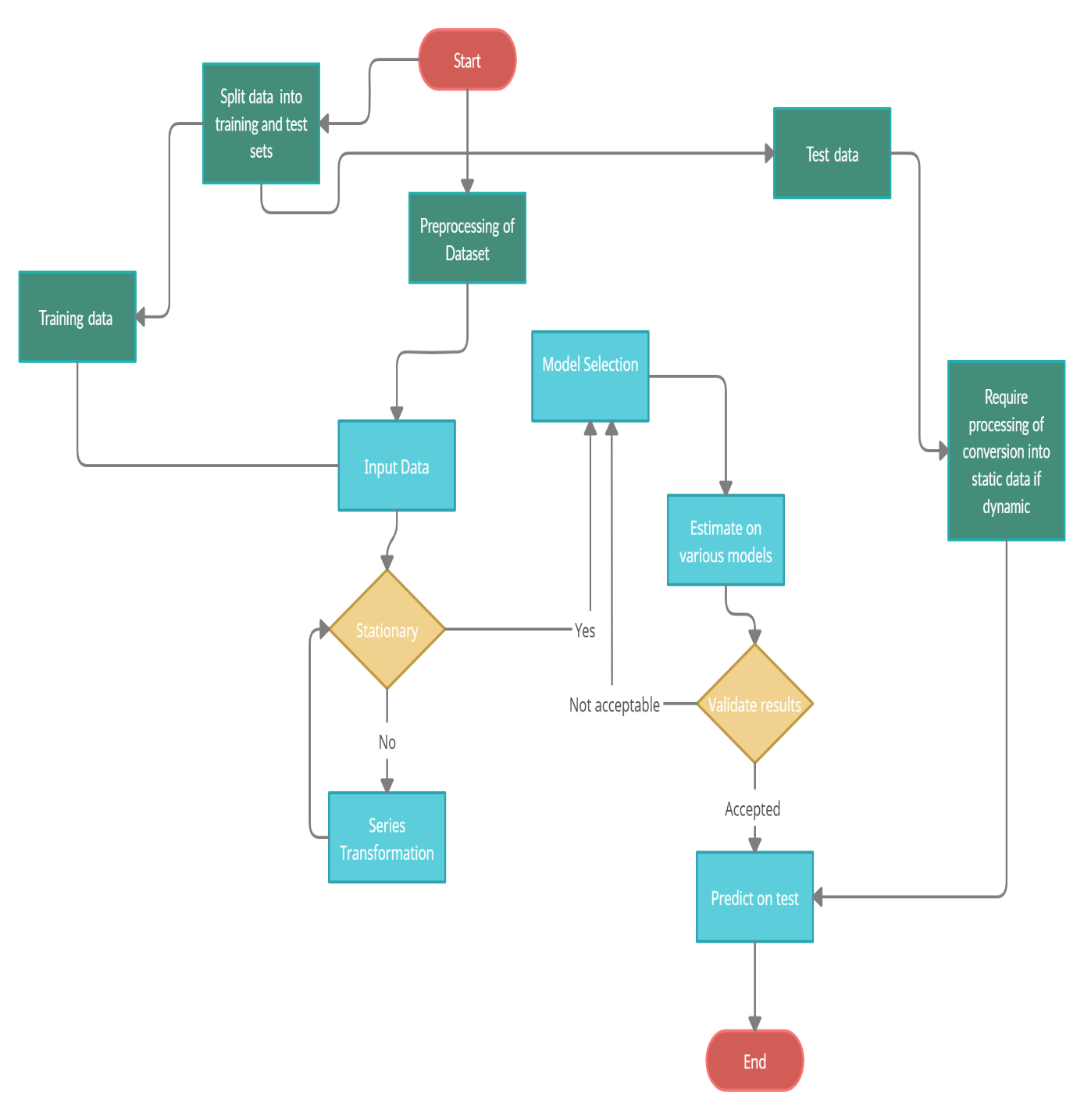
**Model used for time series data is**

**ARIMA model**

Auto Regression Integrated Moving Average is a forecasting algorithm based on the idea that the information in the past values of the time series can alone be used to predict the future values.

The basic idea is to predict value of dependent variable according to time

**FLOWCHART**



**OUTCOME OF THE PROJECT**

* As the outcome of our project, we will be **writing Research Paper** on the project of Profit Prediction

**REFERENCES**

[1] Brynjolfsson E., Hitt L.M., and Kim H.H. Strength in numbers: How does

data-driven

[2] Davenport T.H., and Patil D.J. Data scientist: the sexiest job of the 21st

century. Harv Bus Rev, Oct 2012

[3] Hays C. L. What they know about you. N Y Times, Nov. 14, 2004

Breiman, Leo (2001). "Statistical Modeling: The Two Cultures Statistical

[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C.,

and Wirth, R. (2000). CRISP-DM 1.0. CRISP-DM Consortium, 76, 3.

[5] Donoho, D. (2017) “50 Years of Data Science”, Journal of Computational

and Graphical Statistics, 26, 4, 745-766

decision-making affect firm performance? Working paper, 2011.

[6] Tambe P. Big data know-how and business value. Working paper, NYU

Stern School of Business, NY, New York, 2012.

[7] Kearney V, Haaf S, Sudhyadhom A, Valdes G, Solberg TD. unsupervised

[8] Samuel AL. Some studies in machine learning using the game of checkers.

IBM J Res Dev. 1959;3:210–229.

[9] Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional

dose distributions for external beam radiotherapy. Med Phys.

2016;43:378–387

[10] Valdes G, et al. Clinical decision support of radiotherapy treatment

data-driven machine learning strategy for patient-specific dosimetric

decision making. Radiother Oncol. 2017;125:392–397.

[11] Kearney V, Haaf S, Sudhyadhom A, Valdes G, Solberg TD. unsupervised

[12] Samuel AL. Some studies in machine learning using the game of checkers.

IBM J Res Dev. 1959;3:210–229.

[13] Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional

dose distributions for external beam radiotherapy. Med Phys.

2016;43:378–387

[14] Valdes G, et al. Clinical decision support of radiotherapy treatment

data-driven machine learning strategy for patient-specific dosimetric

decision making. Radiother Oncol. 2017;125:392–397.

[15] Łukasz Kaiser and Ilya Sutskever. Neural GPUs learn algorithms. In International

Conference on Learning Representations (ICLR), 2016

[16] Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser.

Multi-task sequence to sequence learning.

[17] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural

computation, 9(8):1735–1780, 1997.

[18] Oleksii Kuchaiev and Boris Ginsburg. Factorization tricks for LSTM networks.

arXiv preprint arXiv:1703.10722, 2017

[19] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang,

Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding.

[20] Simonyan et al. (2013): "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps".

[21] Springenberg et al. (2015): "Striving for Simplicity - The All Convolutional Net".

[22] Aktekin, T., Polson, N. G., Soyer, R., 2018. Sequential Bayesian analysis of

multivariate count data.

[23] Arunraj, N. S., Ahrens, D., 2015. A hybrid seasonal autoregressive integrated

moving average and quantile

regression for daily food sales forecasting. International Journal of Production

Economics 170, 321–335. 2

[24]. Goldstein, M., Wooff, D. A., 2007. Bayes Linear Statistics: Theory and Methods.

[25]. Willemain, T. R., Smart, C. N., Schwarz, H. F., 2004. A new approach to

forecasting intermittent demand

for service parts inventories. International Journal of Forecasting 20, 375–387. 2

[26] Bohdan M. Pavlyshenko (2018),” Machine Learning models for sales time series

forecasting”

[27] Cui, G., Lui, H. and Guo, X.(2012), “The effect of online consumer reviews on

new product sales”, International Journal of Electronics , available

at: http://www.tandfonline.com/doi/abs/10.2753/JEC1086- 4415170102