A FRAMEWORK USING MACHINE LEARNING TO PREDICT BEHAVIOUR OF CUSTOMERS WITH BUSINESS INTELLIGENCE

Srujan Vannala Ph.D. Research Scholar Department of Computer Science & Engineering, Chaitanya Deemed To Be University, KishanPura, Hanamkonda, Telangana, 506001.
Prof G.Shannkar Lingam Department Of Computer Science & Engineering, Chaitanya Deemed To Be University, KishanPura, Hanamkonda, Telangana, 506001.

Abstract:
Investments in startups can be particularly risky because of their high failure rate. When investing in businesses, individuals like angel investors and venture capitalists run a very significant risk. Using supervised machine learning techniques, we seek to identify the critical characteristics that contribute to startup success in order to help startup investors make judgments.

Key words: Machine learning, Crunch data, Algorithm design, Random Forest

INTRODUCTION
Finding a productive strategy to use the high dimensional data that is available and accessible to them in a substantial and notable way is one of the most difficult problems for an inventive commercial organization. Machine learning techniques have recently been used to predict a set of potential events in any domain. A variety of data mining techniques are successfully combined with new data manipulation tools. Effectively utilizing both supervised learning and unsupervised learning approaches, the prediction process is carried out. In this suggested paradigm, the phenomenon of machine learning is utilized.

The crucial intersection of knowledge discovery and business intelligence is data mining.Many times , the prospective uses of the data that websites provide are not investigated.The future behaviour of p eople is predicted by using their historical data.computer learning.Procedures is used to determine w hich client group will be successful in achieving the desired outcome.

The literature underlying business intelligence should be understood before diving into the suggested model (BI).The gathering of data from multiple sources, its analysis, and its transformation are the t hree main stages of BI.To ensure better decisionmaking, BI is the main goal.A seasonal offer could p ersuade clients to buy if they are unsatisfied with the result they received despite it being attractive.

According to recent research, more people are using smart devices to browse the internet for jobs, shopping, payments, education, exams, business, data, information, entertainment, and more. Therefore, it is crucial for the company to understand what its customers are buying. The major goal of this effort is to foresee and anticipate customer behavior in the online purchasing paradigm. As a result, the suggested machine learning framework performs better thanks to its efficient functionality.

Creating a model is done by incorporating the following key steps: An automated
Model is designed by using data mining techniques and algorithms. The fine-tuned model
Gets the training datasets and gives the predictive analytical results automatically.
The following steps are involved:
- ❖ Gather appropriate data
- ❖ Prepare and transform data
- ❖ Prepare and transform data
- ❖ Build a machine learning algorithm
- ❖ Train, Test and Re-evaluate the model

LITERATURE REVIEW
Laudon and others [63]. Customers compare their options for items or services once they have appropriate information. When conducting a search, they may look for customer reviews or

Client feedback. They research and identify the greatest company or brand that fits them the best. Effective site management and the business's organizational structure are crucial elements that unquestionably affect customers' mindsets. People are therefore preoccupied with making purchases and managing online buying.

Maignan et al. [73] because it helps clients to consider multiple options. Deal administrations and information quality are the most crucial factors that drive customers to choose the products they should buy or which merchant they should buy from during the buying stage of a range of products.

Seock and others [94]. The information in the sites may have a few errors; therefore there is a source danger throughout the research and evaluation phases. Some websites demand registration from users before allowing them to search the site. As a result, consumers also face the risk of information security in addition to product risk.
After making an online purchase, consumers will more closely monitor their post-buy behavior. Customers occasionally have a 27

They may have a problem or concern with the item, or they may need to alter or repair the thing they bought. Return and trade administrations are thus more crucial at this point, according to Liang and Lai [66].

METHODLOGY:
Tuned Advanced Decision Tree Forecasting Model: It analyses the customer's previous data and compares it to the goods being searched for right now. Age, income, and purchasing habits are considered as predictive factors. Figure shows the suggested Machine Learning Framework model.
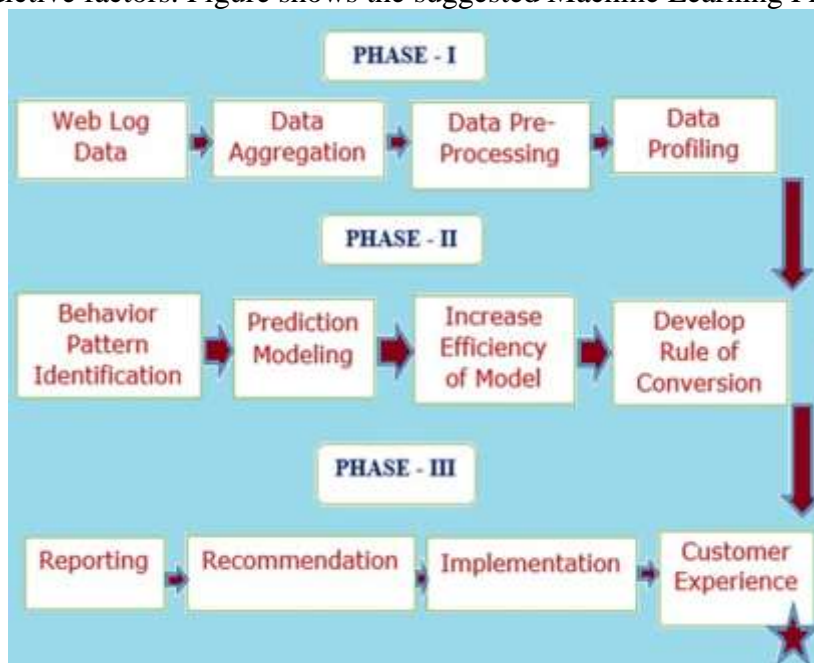

Methods:
Regression Analysis: Many different types of regression analysis are utilized as prototyping techniques. This technique's primary goal is to establish links between a dependent variable, like sales, and one or more independent variables. Elucidating factors, such as consumer spending, the time of year, interest rates, etc.The method can be applied to get a quick glimpse of the factors influencing a variable that is being studied. In order to help a customer select which areas of their organization to invest in, a successful analysis will pinpoint the problems that best explain the change in the dependent variable. Focus on to increase sales, customer satisfaction, etc. Frequently

employed as a forecasting tool is regression analysis. Regression methods include both simple regression and multi-variety regression.

Cluster Analysis: By taking all similarity criteria into account, the factor analysis task creates the groups. Therefore, a cluster analysis will attempt to partition a larger sample into smaller groups where individuals within each smaller group have similar beliefs or behaviors. Once groups have been formed into clusters, each group's common traits can be analyzed, allowing for the detection of behavior among the various consumer clusters. Regarding the type of data we'll be using, the technique is fairly customizable. Cluster analysis can be used to examine overall conjoint findings, as was already mentioned. However, forming clusters can employ any significant quantitative data from a client database or survey. Producing a good cluster analysis is crucial because it is frequently used for market or customer segmentation.

Random Forest: A supervised learning algorithm is the random forest. An ensemble of decision trees, typically trained using the "bagging" approach, make up the "forest" that it constructs. The bagging method's general premise is that combining learning models improves the end outcome. Random forest has the key benefit of being applicable to both classification and regression problems, which make up the majority of modern machine learning systems.

Classification is regarded as the foundation of machine learning. See what a random forest with two trees would look like below:
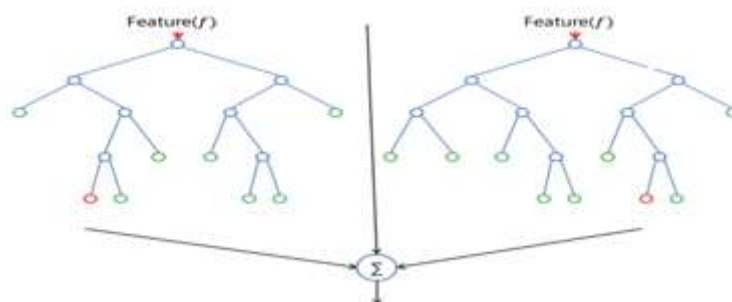
The hyper parameters of a random forest are quite similar to those of a decision tree or a bagging classifier. Fortunately, using the classifier-class of random forest eliminates the requirement to combine a decision tree with a bagging classifier. Using the algorithm's regress or, you may use random forest to handle regression problems as well.

While the trees are developing, the random forest adds more randomness to the model. When dividing a node, it looks for the best feature from a random subset of features rather than the most crucial one. A better model is often produced as a result of the great diversity this causes.

Correlation between the Consumer Behaviors

CONSUMER BEHAVIOUR		
Constant behaviour – people buy the same products	Open-minded behaviour- people are trying new products and begin to buy more	Uncertain behaviour – people don't know what to buy because they have to many options
Past – 1950 Finding balance after two World Wars Past – 1950	1960 – 2000 Exponential growth of World Population 1960 – 2000	2000 – Present Fast evolution of technology 2000 – Present
Political Polling Advertising studies ID it happening?	Surveys Focus groups Interview WHAT is happening?	Observation Online surveys Big Data Neuro-marketing WHY is happening?

Industry, continent and total investment are important features. We received the best result when we used SVM and Random Forest for Multi Class Classification. Received the best result when we used Random Forest for Binomial Classification. For future scope, we would like more data for closed and acquired companies, test model with one-hot encoding, test with other models like Naive Bayes and XG Boost, test with KNN and SVM on Binary Classification Model. Using Crunchbase API, we can also make a real time dashboard and deploy a model so that it can assist investors and founders.

Conclusion:

Predicting a customer's behavior is very tuff task, but by analyzing the market conditions using machine learning algorithms it is possible. Sometimes our predictions may become wrong due to social and economical conditions of the society. For example corona virus affects the common man's life style, in this case prediction will not work, I mean that we cannot get data accurately. But by using regression algorithm we can analyze the customer's behavior by using prototype techniques.

By using cluster analysis we can divide the larger datasets into smaller datasets by this technique we can analyze the customer's behavior and it is very easy to collect the data in cluster analysis technique. According to data sets collected by cluster analysis technique we can train the machine using random forest algorithm. We use supervised algorithms to train machine so it is easy to train the machine because we have already filtered data using cluster analysis hence we may have maximum chances of obtaining accurate results in very complicated conditions . So by following this procedure we can improve the success rate in startups, well established and mid range companies.

References:

1) Akash Rajak and Mahendra Kumar Gupta, "Association rule mining: applications.in various area". Krishna Institute of engineering and technology, DelhiMerruthighway,Ghaziabad, U.P., 2012.
2) Alex Berson, Stephen Smith, Kurt Thearling, "Building data mining applicationsfor customer relationship management CRM". McGraw Hill, 2000.
3) Al-Radaideh. Q, Al-Shawakfa. E, Al- Najjar. M., "Mining Student Data UsingDecision Trees", International Arab Conference on Information Technology(ACIT'2006), and Yarmouk University, 2006.
4) Ana Cristina, Bicharra Garcia, Inhauma Ferraz and Adriana S. Vivacqua,"ArtificialIntelligence for engineering design, analysis and manufacturing fromdata to Knowledge Mining", 2009.
5) Abdul Salam. S.O, Adewole K. B, "In Data mining techniques for knowledgediscovery, Finance institute database", 2004.
6) Abhijit. R, "A source for consumer behaviour analysis in Data Mining Technique",Vivekanand Tarabai Park, 1994.
7) AC Yeo, KA Smith, RJ Willis and M Brooks, "Journal of the operation researchsociety, A mathematical programming approach to optimize insurance premiumpricing within a data mining framework", 2002.
8) Agrawal.R and Srikant.R, "Fast Algorithms for Mining Association Rules in LargeDatabases". In Proceedings of the 20th International Conference on Very LargeData Bases, Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo (Eds.). MorganKaufmann Publishers Inc., San Francisco, CA, USA, pp.487-499, 1993.