International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

# Sales Prediction Using Machine Learning Techniques and Measuring Parameters

**Rajesh Kumar Vishwakarma[1], Dr. Jitendra Sheetlani[2]**

Research Scholar, Department of Computer Applications, SSSUTMS, Sehore, Madhya Pradesh, India[1]

Associate Professor, Department of Computer Applications, SSSUTMS, Sehore, Madhya Pradesh, India[2]

**ABSTRACT***:* In consumer-oriented market it becomes very essential for the early prediction of sales to retain the good profit after selling the products and Accurate short-term sales prediction allows companies to minimize stocked and expired products inside stores and at the same time avoid missing sales. This paper reviews the various existing techniques of data science for sales prediction. In this we also present the literature survey on different model and algorithm used by the various researchers for the prediction of sales. Furthermore, we also discuss some measuring or estimating parameter which measure the accuracy of the prediction of various existing algorithm.

**KEYWORDS***:* Prediction, Data Science, Machine Learning, Data Mining, Forecasting

## I.INTRODUCTION

At presentexceedingly competitive and repetitivelyshifting business environment, the precise and opportuneapproximation of impending sales, also known as sales prediction or sales forecasting, can offer critical knowledge to companies intricated in the manufacturing, wholesale or marketing of products. Short-term propheciesmostly help in manufacture planning and stock management, while long-term predictions can help in business development decision making [1]. Ordering too many leads to waste of products, while ordering too few leads to opportunity loss. Moreover, food consumer demand is constantly fluctuating due to factors such as price, promotions, changing consumer preferences or weather changes [2]. In particular, within the retail and consumer-oriented industries, such as the electronic market or the fashion industry, accurate forecasts are essential. Companies face several challenges regarding accurate forecasts. For instance, they have to place their production plans before exact knowledge about future demands is available. This is required due to the fact that most production plants are located in Asian countries and therefore the time-to-market is longer than the selling period of fashionable products. Therefore, accurate forecasts are crucial because the production of successful products is hardly possible [3]. In addition, other factors, such as changing weather conditions, holidays, public events as well as the general economic situation, can have an impact on future demands [4]. Moreover, fashion items are replaced mostly for the following season; therefore,

there is a huge lack of historical sales data [5]. Summing up, due to short life cycles, high variability in products and demand uncertainties, fashion companies often face high challenges with regard to precise forecasts. Nevertheless, this predictive analysis come with some challenges as the sales are not just dependent on the reviews but many more aspects which may come from the people perspective or may be some changes in a specific brand. As all users have different perspective and does not see a product in the same way. In this research a single year's data is used which is of 9 different cell phone models of OnePlus and likewise the sales of each model are different from each other (price, specifications, camera) and with the launch of every new product having high end finishing and new features increases its sales and decreases the sales of old model. This research is not model specific but brand specific as OnePlus sell cell phones and no other product so only the cell phone review's data is considered. For the prediction of sale various data mining, neural network etc. has been developed but machine learning algorithms which will help in predicting the total sales of these products considering other values of the same review and how efficiently that helps in predicting the sales. This research paper uses data science technique for the sales prediction. Data Science is the mixture of various algorithms, tools, data inference, various machine learning principle and technology in order to solve different analytics complex problems. The main aim of data science is to find the useful insights from the data and the main thing is to mine the raw data and discover the various trends and complex behavior of data. Data science is just like taking out insights from the data that can help the various companies to make better business-related decisions that can give the company a very nice boost. [6] Data Science involves using automated methods to analyze massive amounts of data and to extract knowledge from them. This paper presents literature survey of various data science techniques and earlier

work done in the field of sales prediction. The organization of rest of the paper is as done as follows: Section 2 briefly discusses the work done by the various researchers in sales prediction. Section 3 presents the machine learning techniques and algorithm developed for the implementation of sales prediction. In section 4 explained the measuring parameters of accuracy prediction of sales. And last section presents the overall conclusion of the work and their future aspects.

## II.RELATED WORK

***Tarallo et al. (2019)***presented the benefits of Machine Learning in sales forecasting for short shelf-life and highly-perishable products, as it surpasses the accuracy level of traditional statistical techniques and, as a result, improves inventory balancing throughout the chain, reducing stockout rates at points of sale, improving availability to consumers and increasing profitability.[7]

***Beheraand Nain (2019)***proposed a predictive model usingXgboost technique for predicting the sales of a company like Big Martand found that the model produces better performance as compared toexisting models. A comparative analysis of the model with others interms performance metrics. Experiments support that our technique produces more accurate prediction compared to than other available techniques like decision trees, ridge regression etc. [8]***Chu et al. (2003)***compared the accuracy of various linear and nonlinear models for forecasting aggregate retail sales.The nonlinear versions of these methods are implemented via neural networksthat are generalized nonlinear functional approximators. Issues of seasonal time series modeling such asdeseasonalization are also investigated. Using multiple cross-validation samples, we find that the nonlinear modelsare able to outperform their linear counterparts in out-of-sample forecasting, and prior seasonal adjustment of the datacan significantly improve forecasting performance of the neural network model. The overall best model is the neuralnetwork built on deseasonalized time series data. While seasonal dummy variables can be useful in developing effectiveregression models for predicting retail sales, the performance of dummy regression models may not be robust.[9]***Kumari Punam et al.(2018)*** prediction of sales of a product from a particular outlet is performed via a two-level approach that produces better predictive performance compared to any of the popular single model predictive learning algorithms. The approach is performed on Big Mart Sales data of the year 2013. Data exploration, data transformation and feature engineering play a vital role in predicting accurate results. The result demonstrated that the two-level statistical approach performed better than a single model approach as the former provided more information that leads to better prediction.[10]***Kilimci et al. (2019)*** an intelligent demand forecasting system is developed. This improved model is based on the analysis and interpretation of the historical data by using different forecasting methods which include time series analysis techniques, support vector regression algorithm, and deep learning models. To the best of our knowledge, this is the first study to blend the deep learning methodology, support vector regression algorithm, and different time series analysis models by a novel decision integration strategy for demand forecasting approach. The other novelty of this work is the adaptation of boosting ensemble strategy to demand forecasting system by implementing a novel

decision integration model.The developed system is applied and tested on real life data obtained from SOK Market in Turkey which operates as a fast-growing company with 6700 stores, 1500 products, and 23 distribution centers. A wide range of comparative and extensive experiments demonstrate that the proposed demand forecasting system exhibits noteworthy results compared to the state-of-art studies. Unlike the state-of-art studies, inclusion of support vector regression, deep learning model, and a novel integration strategy to the proposed forecasting system ensures significant accuracy improvement.[11]***Gharibshah et al. (2020)*** proposed two deep learning-based frameworks, LSTMcp and LSTMip, for userclick prediction and user interest modeling. Our goal is to accurately predict (1) the probability of a user clicking on an Adand (2) the probability of a user clicking a specific type of Ad campaign. To achieve the goal, we collect page informationdisplayed to the users as a temporal sequence and use long short-term memory (LSTM) network to learn features that representsuser interests as latent features. Experiments and comparisons on real-world data show that, compared to existingstatic set-based approaches, considering sequences and temporal variance of user requests results in improvements in userAd response prediction and campaign specific user Ad click prediction.[12]Gurnani et al. (2017) evaluated and compared 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. Training data set contains past sales and supplemental information about drug stores. Accuracy of these models is measured by metrics such as MAE and RMSE. Initially, linear model such as ARIMA has been applied to forecast sales. ARIMA was not able to capture nonlinear patterns precisely, hence nonlinear models such as Neural Network, XGBoost and SVM were used. Nonlinear models performed better than ARIMA and gave low RMSE. Then, to further optimize the performance, composite models were designed using hybrid technique and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

decomposition technique. Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost, Hybrid ARIMA-SVM were used and all of them performed better than their respective individual models. Then, the composite model was designed using STL Decomposition where the decomposed components namely seasonal, trend and remainder components were forecasted by Snaive, ARIMA and XGBoost. STL gave better results than individual and hybrid models. This paper evaluates and analyzes why composite models give better results than an individual model and state that decomposition technique is better than the hybrid technique for this application.[13]

## III.MACHINE LEARNING TECHNIQUES

Machine learning has become one of the main-stays of information technology. With the ever-increasing amount of data becoming available there is good reason to believe that smart data analysis will become even more pervasive as a necessary ingredient for technological purposes.[25] Much of the science of machine learning is to solve those problems and provide good guarantees for the solution. [2]

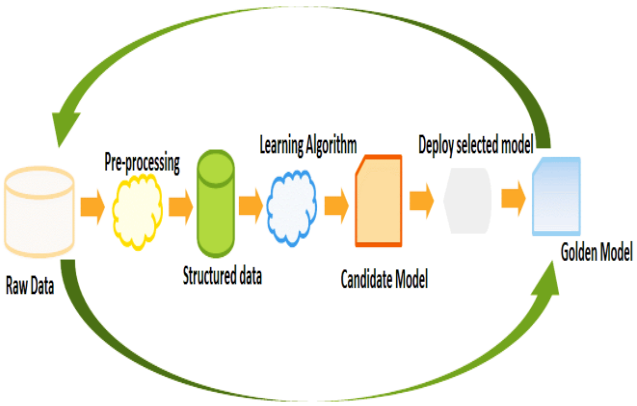


Fig. 1: Process diagram of Machine Learning

In machine learning, various algorithm used for the sales prediction such as logistic regression, random forest, support vector machine, neural network etc. Among various machine learning technique some of them are describing below with their advantages and disadvantages.

### 3.1 Support Vector Machine

SVM (support vector machine) are supervised learning models with associated learning algorithm that analyze data after which they are used for classification. Classification refers to which images are related to which class or data set or set of categories. In machine learning Classification is considered an instance of supervised learning which refers to task of inferring a function from labelled training data. Training data in image retrieval process can be correctly identified images that are put in a particular class. Where each class belong to different category of images. In the SVM training algorithms model is built in which the new examples are assigned to one category class or other. In this model representation of examples in categories are done with clear gaps that are as vast as possible. [24, 26]

The main objective of the SVM machine is to find a particular hyper-plane for which the margin of separation is very high or which can be controlled to be maximized when this condition is met or we can under these circumstances, the decision plane which we take to differentiate between two classes, and then it is called as optimal hyper plane. The Support vectors play an important role in the operation of this class of learning machine as we can define Support vectors as the elements of training data set that would change the position of the dividing hyper-plane in SVM training algorithm if they are removed. As maximummargin hyper-plane and margins for an SVM trained with samples from two classes and these samples on the margin are called as support vectors or we can say that these are data point that lies closest to the decision surface.

Advantages of SVM:

- SVM offers best classification performance on the training data.
- SVM provide more efficiency for pure classification of the future data. It doesn't make any strong assumption on data. It doesn't over fit the data.

Disadvantages of SVM:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

More than one SVM class may accept or all SVM's may reject the data points. In such case data points cannot be classified.

**3.2 Decision Tree**

A decision tree is a graphical representation that makes use of branching methodology to exemplify all possible a=outcomes of a decision. based on certain conditions. In this tree, the internal node represents a test on the attribute, each branch of the tree represents the outcome of the test and the leaf node represents a particular class label means the last decision after all computations. The classification rules are represented through the path from root to the leaf node. It is supervised learning algorithm that is mainly used for classification. They work for both categorical and continuous dependent variables. Decision tree methods construct a model of decisions based on actual values of attributes in the data. Decisions split into tree structure until a prediction decision is made for a given record. Decision trees are trained on data for classification and regression problems. They are often fast, accurate and a big favorite in machine learning. This algorithm splits the data into two or more homogeneous sets. This is done based on most significant attribute to make as distinct groups as possible. For splitting the data into different heterogeneous groups, it uses various techniques like Gini, Information gain, Chi-square, entropy [14-15].

Advantages of Decision Tree [18]

- Clear Visualization: The algorithm is simple to understand, interpret and visualize as the idea is mostly used in our daily lives. Output of a Decision Tree can be easily interpreted by humans.
- Simple and easy to understand: Decision Tree looks like simple if-else statements which are very easy to understand.
- Decision Tree can be used for both classification and regression problems.
- Decision Tree can handle both continuous and categorical variables.
- No feature scaling required: No feature scaling (standardization and normalization) required in case of Decision Tree as it uses rule-based approach instead of distance calculation.
- Handles non-linear parameters efficiently: Nonlinear parameters don't affect the performance of a Decision Tree unlike curve-based algorithms. So, if there is high non-linearity between the independent variables, Decision Trees may outperform as compared to other curve-based algorithms.
- Decision Tree can automatically handle missing values.
- Decision Tree is usually robust to outliers and can handle them automatically.
- Less Training Period: Training period is less as compared to Random Forest because it generates only one tree unlike forest of trees in the Random Forest.

Disadvantages of Decision Tree

- Overfitting: This is the main problem of the Decision Tree. It generally leads to overfitting of the data which ultimately leads to wrong predictions. In order to fit the data (even noisy data), it keeps generating new nodes and ultimately the tree becomes too complex to interpret. In this way, it loses its generalization capabilities. It performs very well on the trained data but starts making a lot of mistakes on the unseen data.
- High variance: As mentioned in point 1, Decision Tree generally leads to the overfitting of data. Due to the overfitting, there are very high chances of high variance in the output which leads to many errors in the final estimation and shows high inaccuracy in the results. In order to achieve zero bias (overfitting), it leads to high variance.
- Unstable: Adding a new data point can lead to re-generation of the overall tree and all nodes need to be recalculated and recreated.
- Affected by noise: Little bit of noise can make it unstable which leads to wrong predictions.
- Not suitable for large datasets: If data size is large, then one single tree may grow complex and lead to overfitting. So in this case, we should use Random Forest instead of a single Decision Tree.

In order to overcome the limitations of the Decision Tree, we should use Random Forest which does not rely on a single tree. It creates a forest of trees and takes the decision based on the vote count. Random Forest is based on bagging method which is one of the Ensemble Learning techniques.

**3.3 Naïve Bayes Classifier**

The Naïve Bayes Classifier is perhaps the simplest machine learning classifier to build, train, and predict with.[16]

Bayes theorem is an extension of the conditional probability. While using Bayes theorem, you use one conditional probability to calculate another one.

Bayes theorem is represented with the following expression:

$$P(A|B)=P(B|A)*P(A)/P(B)$$

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

Here, we calculate the probability of event A, given that event B has occurred. The right-hand side of the equation consists of the probability of event B, given that event A has occurred, multiplied by the ratio of probability of event A to Probability of event B.

Advantages of Naive Bayes

- When assumption of independent predictors holds true, a Naive Bayes classifier performs better as compared to other models.
- Naive Bayes requires a small amount of training data to estimate the test data. So, the training period is less.
- Naive Bayes is also easy to implement.

Disadvantages of Naive Bayes[17]

- Main imitation of Naive Bayes is the assumption of independent predictors. Naive Bayes implicitly assumes that all the attributes are mutually independent. In real life, it is almost impossible that we get a set of predictors which are completely independent.
- If categorical variable has a category in test data set, which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as Zero Frequency.

To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.

### 3.4 Neural Network

Artificial Neural Network (ANN) uses the processing of the brain as a basis to develop algorithms that can be used to model complex patterns and prediction problems. Let's begin by first understanding how our brain processes information: In our brain, there are billions of cells called neurons, which processes information in the form of electric signals. External information/stimuli are received by the dendrites of the neuron, processed in the neuron cell body, converted to an output and passed through the Axon to the next neuron. The next neuron can choose to either accept it or reject it depending on the strength of the signal.[19]

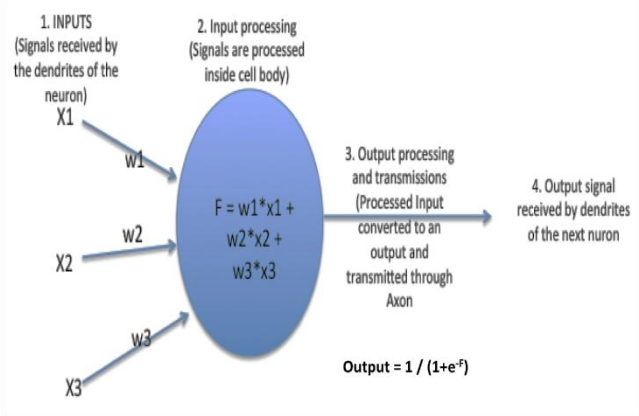


Fig. 2: Artificial Neural Network

Advantages of Artificial Neural Networks ( ANN)[20]

- Storing information on the entire network: Information such as in traditional programming is stored on the entire network, not on a database. The disappearance of a few pieces of information in one place does not restrict the network from functioning.
- The ability to work with inadequate knowledge: After ANN training, the data may produce output even with incomplete information. The lack of performance here depends on the importance of the missing information.
- It has fault tolerance: Corruption of one or more cells of ANN does not prevent it from generating output. This feature makes the networks fault-tolerant.
- Having a distributed memory: For ANN to be able to learn, it is necessary to determine the examples and to teach the network according to the desired output by showing these examples to the network. The network's progress is directly proportional to the selected instances, and if the event can not be shown to the network in all its aspects, the network can produce incorrect output
- Gradual corruption: A network slows over time and undergoes relative degradation. The network problem does not immediately corrode.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

- Ability to train machine: Artificial neural networks learn events and make decisions by commenting on similar events.
- Parallel processing ability: Artificial neural networks have numerical strength that can perform more than one job at the same time.

Disadvantages of Artificial Neural Networks (ANN)

- Hardware dependence: Artificial neural networks require processors with parallel processing power, by their structure. For this reason, the realization of the equipment is dependent.
- Unexplained functioning of the network: This is the most important problem of ANN. When ANN gives a probing solution, it does not give a clue as to why and how. This reduces trust in the network.
- Assurance of proper network structure: There is no specific rule for determining the structure of artificial neural networks. The appropriate network structure is achieved through experience and trial and error.
- The difficulty of showing the problem to the network: ANNs can work with numerical information. Problems have to be translated into numerical values before being introduced to ANN. The display mechanism to be determined here will directly influence the performance of the network. This depends on the user's ability.
- The duration of the network is unknown: The network is reduced to a certain value of the error on the sample means that the training has been completed. This value does not give us optimum results.

**3.5 Logistic Regression**

Logistic Regression is the appropriate regression analysis to conduct when the dependent variable has a binary solution. Similar to all other types of regression systems, Logistic Regression is also a type of predictive regression system. Logistic regression is used to evaluate the relationship between one dependent binary variable and one or more independent variables. It gives discrete outputs ranging between 0 and 1.A simple example of Logistic Regression is: Does calorie intake, weather, and age have any influence on the risk of having a heart attack? The question can have a discrete answer, either “yes” or “no”.Many of the pros and cons of the linear regression model also apply to the logistic regression model. Although Logistic regression is used widely by many people for solving various types of problems, it fails to hold up its performance due to its various limitations and also other predictive models provide better predictive results. [21]

Advantages of Logistic Regression

- The logistic regression model not only acts as a classification model, but also gives you probabilities. This is a big advantage over other models where they can only provide the final classification. Knowing that an instance has a 99% probability for a class compared to 51% makes a big difference. Logistic Regression performs well when the dataset is linearly separable.
- Logistic Regression not only gives a measure of how relevant a predictor (coefficient size) is, but also its direction of association (positive or negative). We see that Logistic regression is easier to implement, interpret and very efficient to train.

Disadvantages of Logistic Regression

- Logistic regression can suffer from complete separation. If there is a feature that would perfectly separate the two classes, the logistic regression model can no longer be trained. This is because the weight for that feature would not converge, because the optimal weight would be infinite. This is really a bit unfortunate, because such a feature is really very useful. But you do not need machine learning if you have a simple rule that separates both classes. The problem of complete separation can be solved by introducing penalization of the weights or defining a prior probability distribution of weights.
- Logistic regression is less prone to overfitting but it can overfit in high dimensional datasets and in that case, regularization techniques should be considered to avoid over-fitting in such scenarios.

**3.6 Random Forest**

Random forests are bagged decision tree models that split on a subset of features on each split. This is a huge mouthful, so let’s break this down by first looking at a single decision tree, then discussing bagged decision trees and finally introduce splitting on a random subset of features. Random forest improves on bagging because it decorrelates the trees with the introduction of splitting on a random subset of features. This means that at each split of the tree, the model considers only a small subset of features rather than all of the features of the model. That is, from the set of available features n, a subset of m features (m=square root of n) are selected at random. This is important so that variance can be averaged away. Consider what would happen if the data set contains a few strong predictors. These predictors will consistently be chosen at the top level of the trees, so we will have very similar structured trees. In other words, the trees would be highly correlated. So in summary of what was stated initially, random forests are bagged decision tree models that split on a subset of features on each split.[22]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

Advantages of Random Forest

- Random Forest is based on the bagging algorithm and uses Ensemble Learning technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way it reduces overfitting problem in decision trees and also reduces the variance and therefore improves the accuracy.
- Random Forest can be used to solve both classification as well as regression problems.
- Random Forest works well with both categorical and continuous variables.
- Random Forest can automatically handle missing values.
- No feature scaling required: No feature scaling (standardization and normalization) required in case of Random Forest as it uses rule based approach instead of distance calculation.
- Handles non-linear parameters efficiently: Non linear parameters don't affect the performance of a Random Forest unlike curve based algorithms. So, if there is high non-linearity between the independent variables, Random Forest may outperform as compared to other curve based algorithms.
- Random Forest can automatically handle missing values.
- Random Forest is usually robust to outliers and can handle them automatically.
- Random Forest algorithm is very stable. Even if a new data point is introduced in the dataset, the overall algorithm is not affected much since the new data may impact one tree, but it is very hard for it to impact all the trees.
- Random Forest is comparatively less impacted by noise.

Disadvantages of Random Forest

- Complexity: Random Forest creates a lot of trees (unlike only one tree in case of decision tree) and combines their outputs. By default, it creates 100 trees in Python sklearn library. To do so, this algorithm requires much more computational power and resources. On the other hand decision tree is simple and does not require so much computational resources.
- Longer Training Period: Random Forest require much more time to train as compared to decision trees as it generates a lot of trees (instead of one tree in case of decision tree) and makes decision on the majority of votes.

## IV.MEASURING PARAMETERS

For the evaluation of sales prediction algorithm, the following experimental protocol is typically used: each time series with the sales of each product is split at the same time point into two parts, the first one chronologically becoming the train set and the latter the test set. The algorithm is then trained and tested sequentially: for each time point of the test set, all data up to the preceding time point are available for training the algorithm. There are various ways to measure the performance of machine learning techniques such as Confusion matrix, Accuracy, Precision, Recall, Specificity, F1 score.[23]

### 4.1 Confusion Matrix

It's just a representation of the above parameters in a matrix format. Better visualization is always good.

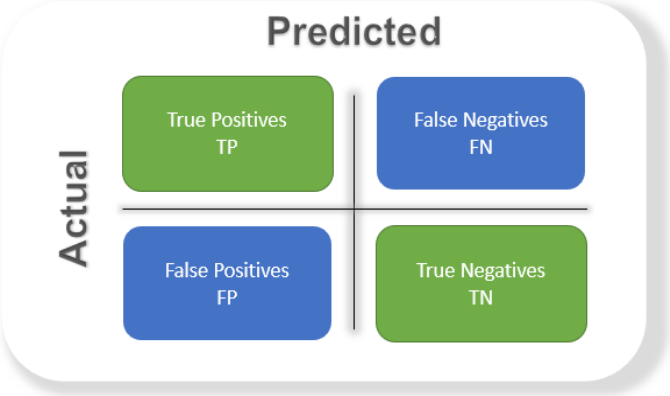


Fig. 3: Confusion Matrix

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 7, July 2020 ||

**4.2 Accuracy**

The most commonly used metric to judge a model and is actually not a clear indicator of the performance. The worse happens when classes are imbalanced.

$$\frac{TP + TN}{TP + FP + TN + FN}$$

**4.3 Precision**

Percentage of positive instances out of the total predicted positive instances. Here denominator is the model prediction done as positive from the whole given dataset. Take it as to find out 'how much the model is right when it says it is right'.

$$\frac{TP}{TP + FP}$$

**4.4 Recall/Sensitivity/True Positive Rate**

Percentage of positive instances out of the total actual positive instances. Therefore denominator (TP + FN) here is the actual number of positive instances present in the dataset. Take it as to find out 'how much extra right ones, the model missed when it showed the right ones'.

$$\frac{TP}{TP + FN}$$

**4.5 Specificity**

Percentage of negative instances out of the total actual negative instances. Therefore denominator (TN + FP) here is the actual number of negative instances present in the dataset. It is similar to recall but the shift is on the negative instances. Like finding out how many healthy patients were not having cancer and were told they don't have cancer. Kind of a measure to see how separate the classes are.

$$\frac{TN}{TN + FP}$$

**4.6 F1 score**

It is the harmonic mean of precision and recall. This takes the contribution of both, so higher the F1 score, the better. See that due to the product in the numerator if one goes low, the final F1 score goes down significantly. So a model does well in F1 score if the positive predicted are actually positives (precision) and doesn't miss out on positives and predicts them negative (recall).

$$\frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * precision * recall}{precision + recall}$$

One drawback is that both precision and recall are given equal importance due to which according to our application we may need one higher than the other and F1 score may not be the exact metric for it.

## V.CONCLUSION

Early prediction of sales product becomes more essential which saves the company from big loss and from the expire of stock of product. It enhances the accuracy of production materials which is on demand in market. According to the prediction they can produce it. For the prediction various machine learning and data mining techniques has been developed. In this work, we present the literature work done for sales prediction and machine learning techniques with measuring parameters. After analysis of these machine learning and data mining techniques it is found that the all techniques have some advantage and disadvantages but it is found that the logistic regression provides more accuracy

than other techniques while decision tree provide less accuracy. In future work need to design the hybrid approach of logistic regression and random forest which will give more accurate results of sales prediction and the analysis of this approach is done on the above parameters.

## REFERENCES

[1] Doganis P, Alexandridis A, Patrinos P, Sarimveis H (2006) Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. J Food Eng 75(2):196–204.
[2] Van der Vorst JGAJ, Beulens AJM, De Wit W, Van Beek P (1998) Supply chain management in food chains: improving performance by reducing uncertainty. Int Trans Oper Res 5:487–499.
[3] Beheshti-Kashi et al., "A survey on retail sales forecasting and prediction in fashion markets", Systems Science & Control Engineering: An Open Access Journal, 2015 Vol. 3, 154–161, http://dx.doi.org/10.1080/21642583.2014.999389.
[4] Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128(2), 470–483. Retrieved May 12, 2014, from http://linkinghub. elsevier.com/retrieve/pii/S0925527310002598.
[5] Thomassey, S. (2014). Intelligent fashion forecasting systems: Models and applications. In T.-M. Choi, C.-L. Hui, & Y. Yu(eds.), pp. 9–27. Retrieved August 19, 2014, from http://link. springer.com/10.1007/978-3-642-39869-8.
[6] https://www.quora.com/What-is-Data-Science-and-Data-Analytics.
[7] Tarallo et al. (2019), "Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research", IFAC PapersOnLine 52-13 (2019) 737–742.
[8] Behera and Nain "A Comparative Study of Big Mart Sales Prediction", 4th International Conference on Computer Vision and Image Processing, 2019.
[9] Chu et al., "A comparative study of linear and nonlinear models for aggregate retail sales forecasting", Int. J. Production Economics 86 (2003) 217–231.
[10] Kumari Punam et al., "A Two-Level Statistical Model for Big Mart Sales Prediction", International Conference on Computing, Power and Communication Technologies (GUCON) Galgotias University, 2018.
[11] Kilimci et al., "An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain", Hindawi Complexity Volume 2019, Article ID 9067367, 15 pages https://doi.org/10.1155/2019/9067367.
[12] Gharibshah et al., "Deep Learning for User Interest and Response Prediction in Online Display Advertising", Data Science and Engineering (2020) 5:12–26 https://doi.org/10.1007/s41019-019-00115-y.
[13] Gurnani et al., "Forecasting of sales by using fusion of Machine Learning techniques", International Conference on Data Management, Analytics and Innovation (ICDMAI) Zeal Education Society, Pune, India, Feb 24-26, 2017.
[14] O. R. . S. P. Breiman L, Friedman J.H., "Classification and regression trees." Belmont CA Wadsworth International group, 1984.
[15] B. C. . U. P.E.T Goff, "Multivariate decision trees: machine learning," no. 19, 1995, pp. 45–47.
[16] [Online]. Available: http://www.statsoft.com/Textbook/Naive-Bayes-Classifier.
[17] http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of-naive.html
[18] http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of-decision.html
[19] Sunpreet Kaur, Sonika Jindal, "A Survey on Machine Learning Algorithms", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2763 Issue 11, Volume 3 (November 2016).
[20] https://intellipaat.com/community/21886/advantages-and-disadvantages-of-neural-networks.
[21] Shah et al., "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification", Augmented Human Research volume 5, Article number: 12 (2020).
[22] Christy Jose ; G Gopakumar, "An Improved Random Forest Algorithm for classification in an imbalanced dataset", URSI Asia-Pacific Radio Science Conference (AP-RASC), 2019. DOI: 10.23919/URSIAP-RASC.2019.8738232.
[23] Salvador García et al., "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power", Information SciencesVolume 180, Issue 10, 15 May 2010, Pages 2044-2064.
[24] Sheeraz Ahmad Peerzada, Jitendra Seethalani, "Machine Learning and Its Implications on Educational Data Base (U-DISE)", Smart Intelligent Computing and Applications, 2020, Springer.
[25] VaseemNaiyer, Jitendra Sheetlani, Harsh Pratap Singh, " Software Quality Prediction Using Machine Learning Application", Smart Intelligent Computing and Applications, 2020, Springer.
[26] Sheeraz Ahmad Peerzada, NeelamadhabPadhy, Jitendra Sheetlani, Gh Hassan, "Predict the Performance of Students and School on Educational Database (U-DISE)", International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020. pp 1-6.