

Restaurant Sales Prediction Using Machine Learning

¹S M Nazmuz Sakib (Orchid- <https://orcid.org/0000-0001-9310-3014>) (sakibpedia@gmail.com)

¹ Graduate of BSc in Business Studies
School of Business And Trade
Pilatusstrasse 6003, 6003 Luzern, Switzerland

¹ Student of BSc in Civil Engineering
Department of Civil Engineering
Sonargaon University
147/I, Green Road, Panthapath, Dhaka

¹ Student of LLB(Hon's)
Department of Law
Dhaka International University
House # 4, Road # 1, Block - F, Dhaka 1213



¹Author Biography

S M Nazmuz Sakib is an eLearning expert and done more than 500 MOOCs or Massive Open Online Courses and experienced as an instructor in sites like UdeMy. He has completed his BSc in Business Studies from School of Business And Trade, Switzerland with CGPA 4 in the scale of 4 and 97.06% grade marks on an average. He is also a certified Google IT Support Professional, Google Data Analytics Professional and IBM Customer Engagement Specialist Professional.

Abstract

In general, the revenue forecast, offer information, and the weather gauge setting will record an accurate estimate of any restaurant's future revenue. The turnover is significantly focused on the need of the customers. Either way, the performance has transformed over the past couple of years with the presentation of huge amounts of information and calculations during the time taken to gain the upper hand. It is fundamental to learn and understand the importance of the information that will be used in any business process. Again, climate forecasting can be done alongside business expectations with the organization.

Keywords: Machine Learning, Sales Predication, Sales, Information Technology, Computer Science

1 Introduction:

The Internet of Things have provided a wide platform to its customer to improve in every field of life. The instant revolution in the technology in from of machine learning, deep learning and artificial intelligence have provide an effective approach to deal with problem while using the modern techniques without any hesitation. Through research it is also concluded with the proper utilization of resources will provide a wide range of opportunity of deal with problems. Restaurants are always considered as a source of food where people come to eat and enjoy. But in the past few years an instantaneous increase in the number or restaurant has created an alarming situation to restaurant to maintain the level of deliciousness of their food, service provided to the customer and other aspects of the restaurants which can be considered as a valuable entity to retain their customer for longer period.

In the past few years, it was noticed that serval restaurants were comparatively creating an enormous amount of profit margin with the help of utilization of resources based on the internet and techniques introduction in order to provide a clearer effective approach to deal with ongoing aspects. The essential goal of this fragment focuses on overseeing machine calculation by following climate and day by day deals of different restaurants. The ideal dynamic section and educated work prospects likewise associate with the general methodologies' dependent on appropriate expectation from Machine Learning Algorithm. The algorithmic model assists with expanding the productivity of legitimate venture that focus on the drawn-out development and improvement perspective in a precise way. A portion of the proposed application helps to foresee in general income of various restaurants chains. Investigative expectation of information and climate anticipating for the eatery become powerful for dealing with the general methodologies fundamentally that can produce development and advancement later on. The significant part of this investigation assists with focusing on limiting human mistakes and play out a few tasks quicker through Machine learning calculation. In general, the income expectation, deals information and climate gauge framework will register precise estimation of any eatery's future income. Different Machine learning calculations further are contemplated focusing on restaurants income stream. For better seeing, high dimensional information improved viewpoint with adjusted methodology of Machine learning rules will be surveyed in this section. Representation of the compelling Machine learning draws near and focusing on the translation can figure

out how to framework has been locked in. Explore the graphical portrayal and follows climate determining report can offer significant help to the restaurants in a powerful way.

Through this research a unique relation between the weather and sales of restaurant can be explored through this research study. The business figure significantly focuses on the clients' necessity. The primary reason for this exploration is to assess the convenience of Machine learning calculation to make a precise deals and climate estimate. Assessment of various information sources that assists with overseeing deals and execute the handy reconciliation to the different information sources. These sorts of information sources likewise help to make deals gauge arrangement utilizing the crude information and recorded information sources. Be that as it may, the best arrangement additionally centers around the relative arranging and usage approaches focusing on the genuine deal esteem. In light of a legitimate concern for the restaurants, an emphasize on the figure of forthcoming days may be more fundamental than to assess the information of coming 2-3 weeks. Moreover, the Machine learning approaches assists with taking out the ideal information and yield incredible outcome that is steady of the general business-related section. The offer of any eatery differs relying upon different factors and climate is one of those significant elements. As the climate changes quickly so following, the clients' necessity interfacing with generally speaking deals of the eatery is huge.

As we are aware of the fact that there are significant chances that human deployed on any respective job can unintentionally make a common error which happens to affect the performance of the system in future during the calculation of revenue at the end of the month or year. One minor error by one individual daily will create 365 errors within the data over a year. This sort of workflow mechanism, based on the minor mistake by every individual, compromises the reliability within the whole infrastructure of the restaurant. However, the fact is quite possible that data cannot be utilized for prediction for the future. For dealing with such worst-case scenario in the past few years machine learning has provided a waste number of algorithms which can be utilized for effective performance in term of accuracy and reliability while performing different tasks. While through this research we are expecting the predict the sales of the restaurants while analyzing the previous data of the restaurant and we will prefer to predict the sales for future so that the management team of the restaurant would be able to create any specific or concrete plan for next year based on the prediction. Another impact aspect of the environment called weather will be related to the restaurant business. The impact of weather on the sales of the restaurant will be evaluated through this research. While for the prediction of sales, the weather forecast will be considered as an essential entity toward the evaluation of the relationship. For effective performance and accurate prediction between different selected relationships, different interactive machine learning algorithms will be implemented in the study to estimates the results. The representation based on the graphical aspect, along with visualization, will be utilized for evaluating the result imposed by the implemented models. At the same time, these representations will also highlight the essential factor, which can strongly impact the sales of the restaurant.

Machine learning is the use of computerized reasoning that gives the capacity to take in naturally and improve as a matter of fact without being clearly customized. Machine learning assists

with taking care of a few sorts of issues; the kind of the issue totally relies upon the dataset. Various methodologies can be taken to take care of the issue. A regulated learning is the Machine learning task, where the information and wanted yield both are given. Either the information is the information or the highlights speaking to the information though, the yield comprises of marks to give the learning premise to expectation of future. The regulated learning is utilized to accomplish the order and relapse task. Order is utilized to sort the information into the classes and the relapse is utilized to figure the numeric yield. The diverse managed models are prepared on the preparation set and make the forecast on the testing set. Various measurements are utilized to assess the presentation of a model. Support vector machine, Naive Bayes, Random Forest calculations are a portion of the instances of managed learning techniques.

Though, machine learning concepts are utilized when there are no names accessible in the information. These calculations discover the example from the information of its own as opposed to learning with names. The unsupervised learning techniques are generally utilized for bunching, where they put the comparative sort of information into similar groups. K-Mean grouping or clustering is the best case of unaided learning strategies. Elbow's technique is utilized to discover the ideal number of groups for unaided learning strategies. On the off chance that the quantities of highlights are more, Principal Component investigation can be utilized for dimensionality decrease of highlights. Along these lines, the significance of understanding the ideas related with the directed versus single learning inside eateries utilizing the Machine learning calculation will have the option to upgrade the general working of the considerable number of procedures attempted in compelling way and as needs.

Restaurants have consistently experienced moderate advancement with regards to utilization of modern techniques based on machine learning and deep learning. Be that as it may, the performance is transforming from the most recent couple of years with the presentation of enormous information and calculations during the time spent increasing upper hands. It is seen that mechanical use in restaurants can by and large think of an answer for the serious market. With the new innovative uses or the new gadgets and applications are being created to beat the serious market patterns. The estimate for deals and climate determining has gotten a lot simpler with the utilization of calculations through the means of machine learning. Nonetheless, there are still contentions going on with improving client involvement in mechanical improvement with the advancement of table top gadgets or the concerned applications that made it simpler for the client to arrange food. Expressed that Tipping Tariff is a significant marvel where clients who are value touchy are bound to change to quicker and speedy administrations from the eateries. In any case, expressed there are developing on the web organizations that are winning greatest benefit through the online application utilizing computerized versatile applications for requesting and conveying the food and administrations. It is seen that 60% of clients are bound to utilize online administrations so as to have food. Thus, it is seen that improving innovations in restaurants and different organizations are additionally animating to improve the deals of eateries. With the online administrations through the Machine learning procedure, the climate doesn't assume a crucial job as the client can arrange online for the nourishments required. During the year 2015, it is seen that the working expense has been confronting difficulties alongside the expense of work and merchandise. As indicated by restaurants are currently

executing new mechanical uses so as to accomplish a serious market. With the utilization of the subjective investigation, the database can be made for the client buying profile. The utilization of information gathering sensors that are being utilized by the restaurants for social affair data of how the client and the representative of the eateries utilize the spot in order to improve the proficiency and furthermore assists with distinguishing the traffic stream by the association. In any case, expressed that it has gotten progressively vital for the eateries for viable information stockpiling and warehousing. Alongside the globalization the proficient mining of huge information can improve the extent of giving upper hands to the separate association for working great in the market.

It is basic to learn and comprehend the significance of information that will be utilized in any business procedure. Be that as it may, with the imaginative procedure of Machine learning and calculations, it has gotten simpler for the organization to sum up the procedure. The various sorts of learning can give better execution by the eateries. With the utilization of machine, learning the deals of the business can be anticipated alongside the other various factors, for example, the climatic variables. Expressed that with the management of learning a precise dataset can be given which are helpful to the organizations in the serious market. Examination can be made as per the set norm and the accomplished standard of the deals and income of the organization for the improvement of the business procedure. To sum up Machine learning, it is essential to build up the model's hypothesis. Which can be responsive to the new information that can be prepared and accomplish precise forecast. Consequently, the precision relies upon the preparation set which can ready to give exact information. As per 2018 characterized Machine learning as over fitting and under fitting which makes the expectation progressively precise and subsequently can be attainable for the future reason. With the expanding independent company, for example, the restaurants, there are sure accessible advantages that are accomplished through machine learning. The individuals are a lot of mindful of the Machine learning procedure, which will assist with recognizing the territories that are required to be progressively improved. Different researcher expressed that with the speculation of the information that are being accumulated for the business advantage and deals reason there are diverse to accomplish physically. Then again, the climate forecast can be made alongside the business expectation with the organization. Different researcher expressed that it is hard to foresee future incentive with creating information in spite of the fact that the business issues can be tackled with the applicable information gathering and setting up the information likewise. To choose the best model for Machine learning is significant and, it mirrors the business benefit. The arbitrary decision of preparing ought to be maintained a strategic distance from so as to beat the opposition. With the speculation of the huge information, it has gotten simpler for the association to accomplish the objective market and procure a benefit from the hierarchical objective. The private company, for example, eateries and online food stores are currently profiting the calculations that are essential for the business. Consequently, this is one of the methodologies taken is seen as helpful for the business association particularly for the little and medium associations which are doing their business utilizing restricted measure of assets and high hazard factors.

2 Literature Review:

Different researchers have contributed toward development process based on machine learning techniques in our daily life routine activities. Sales predication based on weather forecasting with the help of machine learning concepts has being an effective approach to detail with. Machine learning has provided unique mechanism that can be utilized as an effective solution to deal with different real time problem.

2.1 Impact of Weather on sales prediction of Restaurant:

This section previews the impact of weather on the sales prediction of different restaurants according to study conducted by different researchers.

Another effective mechanism that was utilized by (Holmberg and Halldén, 2018) in order to achieve forecasting or sales in restaurant based on the concepts of machine learning is considered as a one of the most appropriate techniques in the domain of restaurants. In the research study three different dataset were utilized for the purpose of experimentation based on three different cities. Whereas the data extraction was performed on the basis of SMHI concepts and data pre-processing process, correlation of features, standardization and normalization are involved in the implementation process. Further root means square error, extreme gradient boosting, model setup, final XG Boost model, LSTM neural network involves the setup of model and at last final LSTM model utilization were part of the implementation process. The most significant highlights while foreseeing the deals are plainly the highlights with respect to date, while the climate highlights has the least effect. This is for the most part valid for all restaurants, obviously there are eateries in which the climate highlights have a bigger effect than in others. The assessment and examination of the different eateries gave a thought of the assortment between the kinds of eateries and the absence of highlights in our models. All restaurants are accepted to be occasion subordinate, yet at different degrees. The occasion information, outer or inner, are relied upon to have a greater effect than the climate information by and large. Whereas evaluation of the implemented model was based on date features, weather features, features importance and selection, summer and weekday dataset.

According to different researches the fact was concluded that preferences of customer are based on season and weather which impact the commodity of sales generated by any restaurant. A study conducted by (Abadi et al., 2016) which evaluate the system of TensorFlow which is considered as machine learning infrastructure based on large scale. The capability of the considered system lies on the evaluation of different parameter in Big dataset without any jurisdictions. The model for execution of TensorFlow flows a specific pattern to evaluate different state and computation of algorithms based on machine learning. The tensor model is capable of executing concurrent on multiple based on a subgraph of overlapping. while the element related to graph for dataflow were operated within different operators Tensor, operations, operations of stateful based on variables and stateful operations based on queues. While the following activities were involved within the methodology in the research study such as execution based on concurrent and partial, execution based on distributed manner and dynamic flow of controls. Cases studies of extensibility of the implemented model was carried out in the following way which can be illustrated as the following optimization and differentiation, training process of large models, tolerance of fault, synchronization based on coordination of replicas. The evaluation was performed through the following techniques

which can be described as following single machine based on benchmark, synchronous microbenchmark replication, classification of images and modelling of language.

The impact created by weather over sales of restaurant were evaluated in a study proposed by (Liu and Ichise, 2017). The study specifically considers the supermarket chains in Japan. The study briefly described the important prospective of encounter and LSTM which were part of the research methodology. The implementation of LSTM in the proposed study clarify the dependence of different factors related to sales over weather variations. While the prediction through LSTM utilize the factors of days, date, weather and sales for the purpose of prediction. The method of LSTM takes input in the takes in the input in the form of sequences. While the main analysis of LSTM was carried out in different layers and further analyzed under different layer for the autoencoding then the final output is obtained in a label. For evaluation of proposed model from 2012 to 2013 data relate to supermarket sales in Japan was utilized. For the comparison of performance based on the factor of accuracy was evaluated as compared to other traditional approaches of machine learning AdaBoost, LR, GBDT and SVM. While further evaluation of the proposed study was based on the factors of score of F-1, recall and precision. Through analysis the researcher concluded that proposed methodology outperforms in different factors as compared to other method and techniques based on machine learning.

According to the business point of view prediction on sales provide an assumption based on previous record of business which can be used for create manifesto, effective business model and strategy which can provide effective results in terms of improvement. Through the research conducted by (Meulstee and Pechenizkiy, 2008) which emphasis on the prediction of sales while highlighting the aspects which were not utilized by any other researcher before. The research highlights the challenges faced during the prediction of food sales while considering the case study of food group company of Sligro. The research tends to implement the modern approach of ensemble learning which emphasized on the integration dynamically for the purpose better classification of different functionalities as per the requirement of the customer. Improvement in the sales through prediction was evaluated through analyzing different factors such as public holidays impact on sales, weather. Feature construction process in proposed study is illustrated in the Figure 02. While the experimentation was based on selection of targeted data and ensemble learning. Experimental analysis was based on the average of MSE, accuracy and absolute error. The researcher concluded that performance of ensemble approach of learning tends to surpass the performance of traditional algorithms utilized for prediction. Through implementation of ensemble technique, the researcher noticed that effective extraction of data over the impact of weather was noticed. The significant factor which was

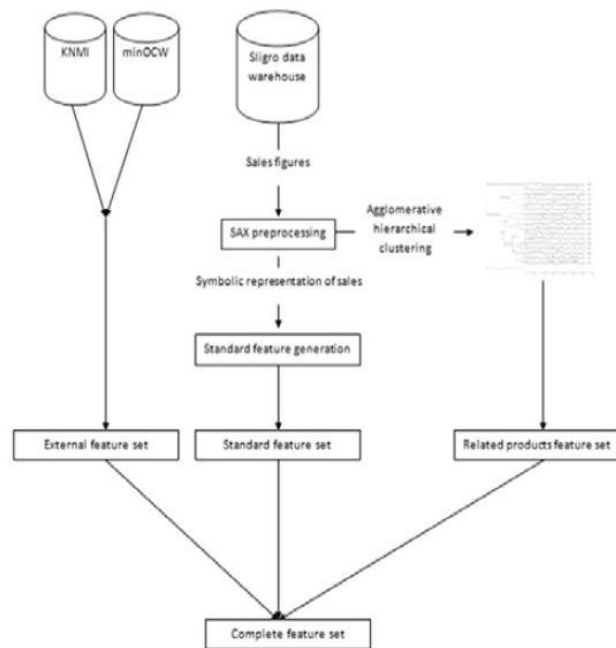


Figure 1 Process of construction of Features by Meulstee et al., 2008

conclude from the research was that appropriate assignment of data related to public holiday and weather produce accurate prediction as compared to traditional approach implemented for this purpose.

2.2 Machine learning Techniques for Prediction:

This section briefly reviewed different machine learning techniques on the basis of supervised and unsupervised learning concepts for prediction.

2.2.1 Unsupervised Techniques

The effective implementation of unsupervised machine learning algorithms provides clear pattern and result when no specific categorization within the dataset exist. The capabilities of unsupervised algorithm rely on the cluster of datasets to evaluated different patterns within the dataset. A study proposed by (Claypo and Jaiyen, 2015) which implement the unsupervised algorithm of clustering through K-mean and selection of features through MRF which evaluates the reviews of restaurant based on Thailand. The study tends to cover the mining of opinion related to millions of reviews provided by customer to different restaurant in Thailand. The study focusses on attitude, opinion, emotions, evaluations and sentiments illustrated within the process of opinion mining in the proposed study through reviews of customer. Through different studies the fact was concluded that improvement in the services and food quality can be enhancement with the help of results concluded from mining of opinions. The study adapted the text processing through splitting the while mechanism into two parts in which the first part the reviews were transformed into token while further the next part involved the removal of those words which doesn't carry any specific meaning. The transformation of text was carried out in two part which were related to creation of keyword lists in the first part. Further the second part involved the creation of a formatted text file based in the format of numbers. The selection of features was carried through implementation based on optimization technique of Markov random field. The research implements the dataset of the trip advisor which consolidate the reviews of restaurant reviews. Further the clustering was performed through FCM, Hierarchical, SOM and K-means based factor of time and accuracy for the purpose of evaluation. Through experimentation the fact was conclude that computability of K-mean clustering surpasses the selection features mechanism of MRF in term of clustering.

2.2.2 Supervised Techniques

According to study proposed by (Te et al., 2018) which implement the machine learning while utilizing the resources of web data services based on different restaurants data for the purpose of increasing the growth of the restaurant. Through this research, the data mining is performed

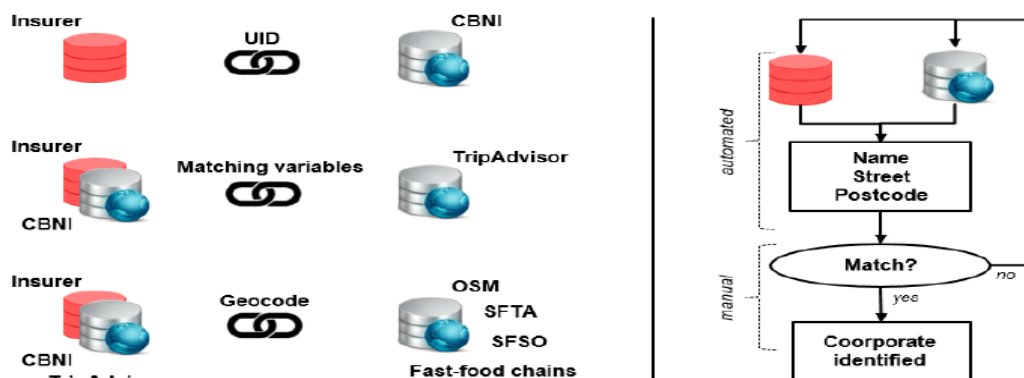


Figure 2 Data linking scheme for sample based on Web adapted by (Te et al., 2018)

on the large-scale data in order to achieve better pattern of policy and factor which may help the restaurant to increase their existing results in term of efficiency and accuracy. In the study the researcher implements the supervised algorithm based on machine learning in order to build pattern between data whereas comparison was considered another important feature of the study. The method involved for collection of data involved insurer data, web data based on central business names & trip advisor, open street map, swiss federal statistical office, swiss federal tax administration and fast food chains. Whereas the creation of label process and pre-processing of data were performed in a way such as growth label creation, input feature creation and construction of the growth model. whereas evaluation of the proposed study was based on the factor such as AUC, accuracy, sensitivity and specificity. Whereas random forest, multi-layer perception and logistic regression methods were utilized in the proposed study for the purpose of evaluation of the proposed model. Figure 1 shows the cooperate approach adapted by research for the linking of corporate data collection over the web samples. The research study was based in Switzerland, the researcher conclude that the results of the experimentation might have different outcomes over different regions.

In domain of business, feedback and ratings are considered as interactive review from customers after utilizing the serves of business. Over the period of time different companies were able to perform effective as compared to its competitors after following the reviews and feedback of customers. In the domain of restaurant, in order to predict rating a study proposed by (Kulkarni, et al., 2019) address the idea by utilizing the concepts of algorithms based on machine learning for the purpose of prediction of rating. The proposed research utilized different machine learning techniques such as Decision Tree, Random Forest, XG Boost, Support Vector Machine, linear regression and ADA Boost in order to predict the rating of restaurants. Whereas evaluation prefer to follow the following parameter in order to utilize further for the process of prediction based on votes, area situated, cuisines, reviews and average cost for two people. Data pre-processing was performed on the data in order to filter or extraction of actual data instead of raw data. During the feature selection analysis following parameters were dropped form the dataset such as location, phone, address, review list and URL. Whereas exploratory data Analysis were further performed on distribution rate of restaurant, approximation of two people cost, rating based on online ordering, rating which deal with respect booking table, restaurant based on top rated and distribution based on online booking and ordering based rate and cost. Through experimentation it was concluded that accuracy rate of prediction of 83 percent was achieved through the utilization of ADA Boost technique.

According to survey conducted by (Tsoumakas, 2019) which tend to reviews different techniques of machine learning for prediction of sales in the domain of restaurant. The research tends to highlight the best approach which might be utilized in order to produce effective prediction based on the flow of data within the dataset. The evaluation of machine learning approaches and techniques of the accurate prediction of food sales are analysed through measurement of granularity within the temporal. The research highlights the perspective of output variables on the basis sales produce by each product while the weight was marked as another important entity toward development. While the evaluation of the experiment was based on the factor of accuracy which was based on the following measures such as MSE, RMSE, RRMSE, MAE, MAPE and MASE.

2.3 Effective Techniques for Prediction:

The goal of every business is to achieve success in term of achieving the profit targets. According to research studies it is concluded that effective policy with the time should be adapted in order to achieve better and efficient outcomes. In the domain of restaurants business, the company has to follow different unique strategies in order to achieve better result in the shape of achieving the profit margins. The modern technology means it is concluded from different research that evaluation based on existing data of a successful business will provide different specific factor which were adapted by business in order to achieve successful. To deal with such scenario according to study proposed by (Santhana Lakshmi and Bavishna, 2020) which tend to analyse the existing data of restaurant using the modern technology in order to find out those factors which are actually responsible for increasing the sales so implementation of those factors might possibly boost the income of the restaurant. The study was carried under two different modules in order to figure out those factors which are responsible for increasing while evaluating through prediction modulation and analysis modulation. This study explores the monitoring of the restaurant system based on online system of delivery utilized for customer service along with those customers who directly approach restaurant for food. For the purpose of evaluation of the dataset Power BI and R Tool were utilized in order perform linear regression and other analysis on the dataset. Whereas following parameters were involved within the process of linear regression date, in dinner, take away, bulk order, diner, ambience, service, facilities, taste, Zomato, online, Zomato rating, swiggy, items, total, swiggy rating and diner rating were part of the evaluation of analysis and prediction modules. Through analysis it was found out that swiggy sales were the main reason behind success of the restaurant sales.

Classification of big data on the basis of sentiment analysis through implementation of apache spark were evaluated in the study conducted by (Al-Saqqa et al., 2018). The study tends to highlight the correlation of different factor which are involved within the process of prediction. Evaluation the different corelated aspects on the basis of sentimental analysis was considered as one the significant factor of this conducted research. The proposed research not only briefly elaborated the significance of analysis based on sentiments, Apache Spark and big data. For the purpose of evaluation, the classifiers of SVM, LR and Naïve bayes were implemented under the command of environment setup in Spark. The research implements the dataset of reviews though amazon which were evaluated in the way such that over 2 M samples were evaluated within testing phase and over 1.8 M samples were evaluated within training phase. The pre-processing of dataset was carried out for cleansing of classifier and further preparation of classifiers. In the domain of the pre-processing certain review will null values were removed from the dataset and further separation based on characters were carried through implementation of tokenization within the research. Further the extraction feature in the research was based on conversion of textual data into vector of features. The researcher concluded that performance based on the factor of accuracy was provided through the implementation of SVM.

According to (İşlek and Öğüdücü, 2015) the forecasting process of demands related to a business company which tend to provide complete overview of product quality which meant to be purchased by customer. The paper focus on the utilization of effective model for forecasting in order to remove limitation during the whole process while achieving higher accuracy in terms of estimation or prediction. The adaptation of clustering of factor was

perform through graph of bipartite process of clustering. Further the implementation of machine learning approach of Bayesian Network along with average moving model for the phase of hybrid forecasting. The involvement of direct supply chain, extended supply chain and ultimate supply chain within any organization was highlighted through the proposed research. The representation through the means of nodes and edge were handled through bipartite clustering of graph. For the assertions of conditional factor based on independence were represented graphically through the network of Bayesian. The methodology of the paper was conducted in the following manner such as construction of dataset, calculation of average moving values, construction of graph through bipartite, clustering through bipartite, implementation of machine learning and then the final result in the form of forecasting were achieved. For evaluation purpose dataset of companies of turkey were considered over the period of 2011 to 2013. The evaluation was based on the factor of MAPE for measurement of rate of error. The researcher concluded that effective improvement based on performance during forecasting was achieved through this research.

According to research conducted by (Boyapati and Mummidi, 2020) which evaluated different machine learning algorithms for the precise prediction. while the researcher tends to find those factors, which were responsible for effective increase in sales. The research performed the prediction through the implementation of simple linear regression, gradient boosting regression, support vector machine, random forest regression along with other supervised learning, unsupervised and reinforcement learning. The environment for experiment was based following tools and software for analysis such as python, NumPy, Pandas, Matplotlib, Sklearn, Seaborn. For the purpose of evaluation, a dataset containing 12 different attributes covering over 8523 instances was part of the research study. The study considered the following factors for purpose of finding correlation between data such as identifier, weight, Fat content, visibility, type, MRP related to items while as corresponding to outlet the following factors were evaluated such as type, sales, location, identifier and year in which the outlet is established. Data pre-processing process was carried out through encoding of values in categories and cross validation through stratified K-fold. While the evaluation of performance was based on the following factors such as score of accuracy, maximum error, absolute error based on mean values. For comparison of different algorithm, the evaluated was performed through calculation of average score of accuracy, average of absolute error mean value and average of maximum error. Through analysis the fact was concluded that the implementation of regression through random forest outcomes on the basis of different evaluation factor surpass other implement machine learning techniques and method.

2.4 Challenges during Sales prediction

Another important aspect that can be considered as an important entity toward development of the restaurants can be illustrated on the basis of reviews. They are considered as one of the most efficient and effective tools to evaluate the performance of any restaurants whereas these tools can also utilize as an effective mechanism to introduce different policy based on the feedback of different customers. In this specific domain issues were further addressed by (Hossain et al., 2017) which utilized the review of different customers and implement the machine learning techniques in order to achieve better performance based on sales of restaurant. For the purpose of experimentation, the research was based on the sample dataset embedded the review of people as Priyo dataset. Whereas over 50 selected restaurants were part of the dataset. Whereas the research utilized the tools of natural language toolkit for the

process of pre-processing. Whereas the creating features involved review selection, tokenizing and part of speech tagging, selecting keyword and building classification model. whereas process of evaluation involved four different algorithms such as logistic regression, support vector machine, k nearest neighbour, and multinomial naïve Bayes, Whereas performance of the implemented model was based on the factor of Accuracy, TPR, TNR and AUC. Through this research the researcher concluded that the impact of review will provide unique mechanism to different restaurants to identify their market statistics.

The location of business point is also considered as a crucial entity toward business growth and development in a very short span of time. According to different research it was concluded that traffic situations around business point creates extra opportunities for business awareness in common people. Such scenarios were covered by the research proposed by (Abrishami et al., 2017) which review the traffic situation around different business including restaurants in order to find out the result which are occurring due to this factor. The research covers the data of 100 business over the period of 1 year based on the traffic situation around the business. For the collection of AP Mac, Client Mac, Seen Epoch and RSSI were utilized in the process along with data processing was also based on these mentioned factors. For the evaluation of error and measurement of those error during forecasting MAE, RMSE and MAPE were utilized. Whereas the prediction model was based on the following factors such as function learning (based on weather, holidays, special event and location). While the evaluation of the proposed research the business was categorized into 6 different categories such as gym, coffee shop, restaurants, bar and barbershop. Whereas evaluation of the dataset was performed on different machine learning approaches such as utilization of python package sklearn in order to utilize the service of regression based on random forest, Libsvm in order to utilize the services of regression based on support vector machine along with predictor based on cloud services of Google. The experimentation process was divided into five different stages. Through experiment result it was concluded that's accuracy and performance of SVM in order to perform evaluation was effective and efficient as compared to other utilized techniques.

The study proposed by (Aishwarya et al., 2020) address a critical issue related to restaurant management. The research emphasized on the technology means of which utilizes the services of learning based on concept of machine for the prediction of food demand. The idea behind this proposed research was to address and manage the food services with the passage of time and review them as per prediction. The main reason for conducting this research was to provide an effective and accurate mechanism which allows the management to plan their buying or export of ingredients in a mean while time. Whereas in the domain of restaurant we are pretty much aware of the most of the item have life based on short term period. The research also utilized the concepts of demand forecasting which emphasize on the estimation as per the demand of customer based on the data based in history. This research targets the quantity of ingredients based on the orders utilized against them while analysing them using the techniques and algorithms based on machine leaning. This research utilizes the services of linear regression, decision forest regression, Bayesian linear regression, support vector machine and XG Boost for the purpose of dealing with forecasting method. While the statistical method based on dataset was carried through stepwise method in the proposed study. Whereas forecasting based on the number of customers in the proposed research was carried in the following cases such as on track variable along with forecasting was performed via the following parameters such as number order, homepage featured, email

for promotion, operation area, city code and region code. Through evaluation it was concluded that every utilized algorithm for the purpose of prediction shows similar result based on performance for forecasting.

Different issues are associated with the implementation of algorithms based on machine learning. The uncertainty within the signals will always overcome the financial aspects according to research conducted by (Chen et al., 2018). The appropriate implementation of learning techniques based on machine learning will provide effective mechanism for encountering those strategies which might turn out to be disaster for business or damage the performance of business based on facts and figures. For establishment of system which is based on machine learning concepts which can further utilized in the management of financial aspect of any restaurant consumes a bit of time frame for the purpose of effective development of the ideal system. Different researcher has noticed that still some effective factor should be discovered in the domain of business in order to adapt the modern means of technology for effective performance according to (Francis and Kusiak, 2017). While the issues related to accuracy within the workflow of restaurant were notified as a serious lack of the system. Through effective implementation of algorithms based on artificial intelligence the accuracy on the any proposed system can be achieved.

We are familiar with the fact that for performing any specific task there are various number of ways evaluated on the basis of performance factor. Such a study which evaluated the performance of different techniques for prediction of sales and reviews were conducted by (Lasek et al., 2016). The study focusses on the method and techniques utilized for prediction of demand of customer and sales of restaurant. Along with evaluation of other proposed model the research tends to highlight the important aspect of forecasting of demand. The effective implementation of any prediction model can be possible through an effective management system of revenue which can adjust the prices of food as per the demand of the customer. For the purpose of prediction following parameter were utilized within the project such as time, weather, holiday, promotions, events, historical data, indicator of macroeconomics, competitive issues, websites, type of location, location of demographics. The research tends to highlight the performance of the following algorithms such as multiple regression, passion regression, model of Box-Jenkins or known as ARIMA, Holt winter and model based exponential smoothing, artificial neural network, Bayesian network and other hybrid models. The research highlights some associated rules of analysis based on the market basket. While the input, output, advantages and disadvantages related to above illustrated algorithm and techniques are part of the research paper.

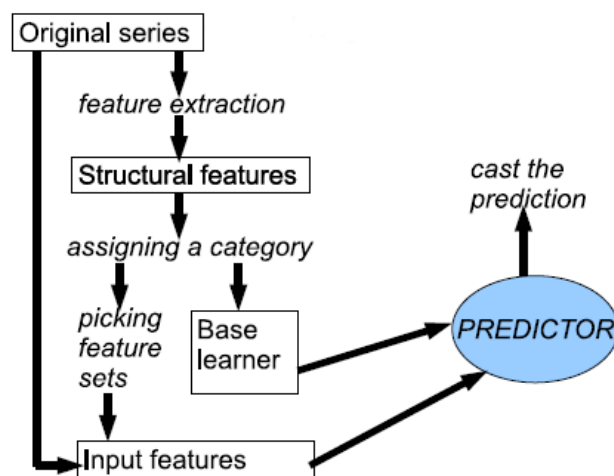


Figure 3 Online Operation in CAPA by Zliobaite et al., 2009

In order to maintain the profitability of business and management of stock in the domain of business the company should undergo for predictions on time basic in order to evaluate statistics based on accuracy. The research study conducted by (Žliobaite et al., 2009) tend to predict the sales based on content aware food. The approach of CAPA tend to selects the algorithm for predictions based on the historical sales properties presented in structural way. While the operation of online activity and offline activity were implemented in the CAPA through the means of decision support. Figure 3 illustrate the CAPA online operation. While the categorization of learning was based on the factors such as bottom up categorization which embedded the methodology of validation & training, and top down categorization. The data reside in the external and internal system are utilized in the process of forecasting which are furthermore utilized within features input space. For the purpose of evaluation Sligro group food industry case study was implemented which contain a data of six thousand products and 40 different outlets. The setup of experiments was based on the factor of bases predictor selection, categorization of learning, accuracies based on prediction. The researcher concluded that the most difficult encountered throughout the whole research was proper categorization for extracting accurate results.

3 Analysis & Exploration of Data:

3.1 Overview of Analysis:

In the past few year, the food delivery through online services is effectively adapted by customer. The preference of customer is changing form eating food at restaurant to ordering food through online services. The dataset comprises the data of restaurant based in Bangalore India for the purpose of analysis and evaluation. while the exploration of implemented model was based on the analysis of two different datasets. The dataset of order wise sales summary or OSS contain 45 different number of features and 33,401 number of rows embedded within it. The other dataset based on customer data of restaurant contain 10 different features and 28,160 number of rows were part of the dataset. For effective prediction specific features were cleaned from the dataset which was utilized though the implementation of data cleaning process in the dataset.

As a result of data cleaning, specifically 9 different features were extracted form overall 45 features of OSS dataset. Form the CDR dataset only 7 different features were extracted from the overall features. Through Table 1 the features selected form OSS dataset is illustrated. While the illustrated of cleaned features form CDR dataset are represented through Table 2.

Features No	Features Name in OSS Dataset
1	Cancellation
2	Delivery Time
3	Grand Total
4	Order Date
5	Order Source
6	Packaging Charge
7	Payment Mode
8	Pre-order

9	Restaurant Discount
---	---------------------

Table 1 Data Clean Features form OSS Dataset

Features No	Features Name in CDR Dataset
1	Phone No
2	Customer Name
3	Last transaction
4	Last transaction Days
5	Total orders
6	Revenue
7	Average Basket Size

Table 2 Data Clean Features form CDR Dataset

The dataset of order wise sales summary was based on the information of sales related to the time period of year 2017 to 2018 comprises the details of sales generated while utilizing the online services. Different analysis was performed on the dataset in order to evaluate different trend and factor within the dataset while utilizing the concepts of machine learning. The correlation between number of customers utilizing different online food service is analysed under machine learning approach. The results concluded from the analysis is illustrated in Figure 4 which clear shows what significant number of users uses the service of Zomato and Swiggy. The effective services provided by Swiggy and Zomato are significantly trusted and utilized by customer all over the region.

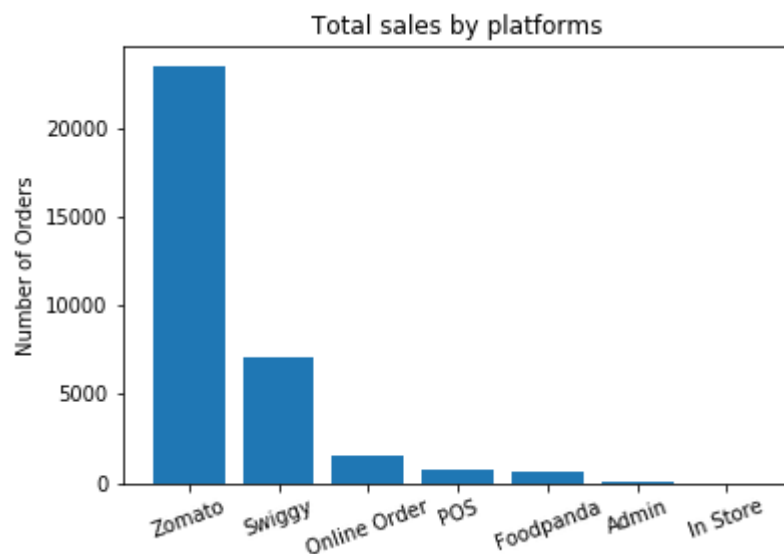


Figure 4 Order Placed by Customer on Different Platforms

This study further evaluates another interesting fact which emphasized on the booking of pre-order for food by customer. This sort of feature is provided to customer to manage the food service for different schedule meetings and sittings. Figure 5 illustrated the usage of pre-order feature by different customers. Through analysis the fact was concluded that very low percentage of customer prefer to go for pre-order service. While most of the customers prefer not to go for the pre-ordering service because of different restriction are associated with the pre-order features of Swiggy, Zomato and other online services.



Figure 5 Usage of Pre-order Service by Customers

Online services have provided interactive platform to customers in terms of service, reliability and effective domain for helping customer. The mode of payment is also considered effective entity toward customer care and service. Figure 6 illustrated the different payment mode utilization for using the resources of any online food service. Through analysis the fact was concluded that most of the payment were performed through online banking services. Through analysis the fact was find out that 39 minutes average time for delivery is required for the delivery of food to the end customer.

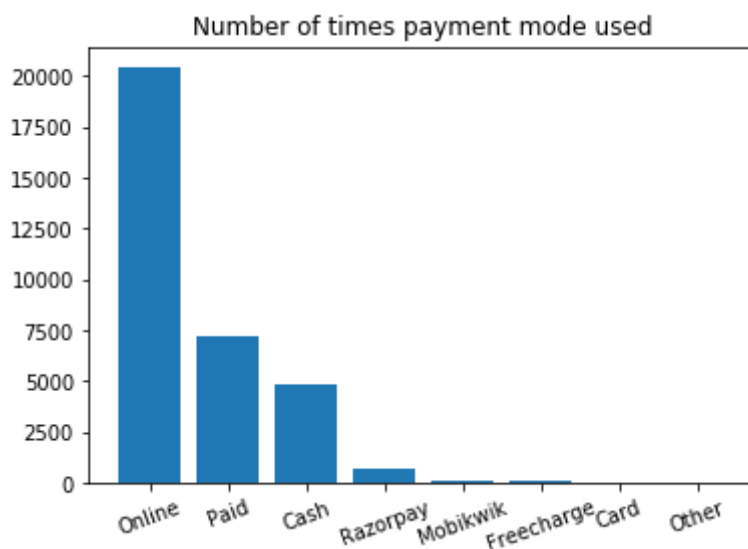


Figure 6 Usage of different mode of Payments

3.2 Impact of Weather on orders and Daily Sales:

The fact can be denied or prove wrong that condition of weather influences the orders and daily sales of restaurant. According to different research the fact was concluded that seasonal changes impose a significant variation within the order and daily sales of restaurant. While different factor of weather is also associate with human psychology which let them to order food through different restaurants. In order to perfectly organize the demand of customer according to weather different analysis should be conducted in order to carry some precautions while dealing with customer in different season changes. Through different research the fact was concluded that one the most significant part contribution is naturally provided by the factor of weather in order to improve the performance of restaurant.

The analysis based on the relation between impact of weather on sales is illustrated in Figure 7. On the other hand, other analysis, the relation between impact of weather on revenue generated on those sales is illustrated in Figure 8. Both the analysis will provide effective understanding over the impact of weather on sales of restaurant.

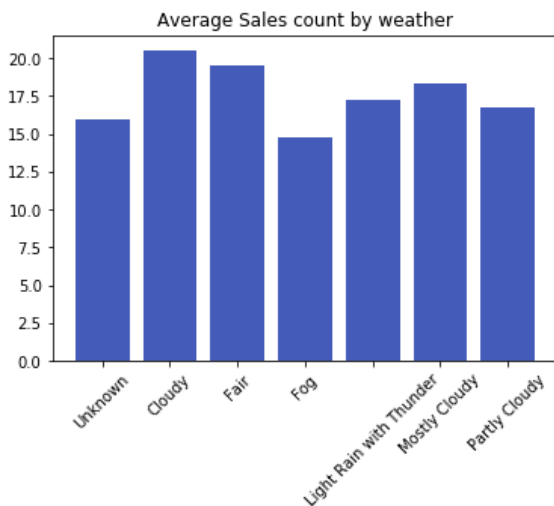


Figure 7 Impact of Weather on Count of Sales in Average

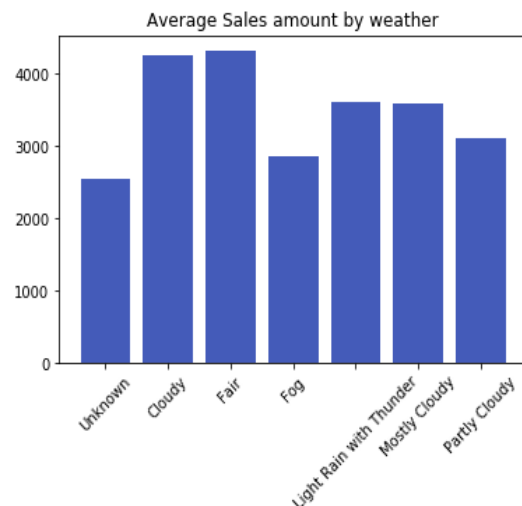


Figure 8 Impact of Weather on Cost of Sales in Average

Through analysis the fact was find out that during cloudy and fairy weather most count of sales was noticed. While the least number count of sales was noticed during foggy weather.

3.3 Impact of Public Holidays and Weekend on Orders and Daily Sales:

The fluctuation within sales and order of restaurant shifts from season to season. The impact of public holidays can be easily understood after conducting different analysis on the relevant relation. The human psychology prefers to find calmness on weekend after having a rough week they prefer to order food online to spend some quality time with their family and find pleasure in life. In order to completely understand the dept of the situation analysis was conducted between total number of sales against the number of weeks in Figure 9.

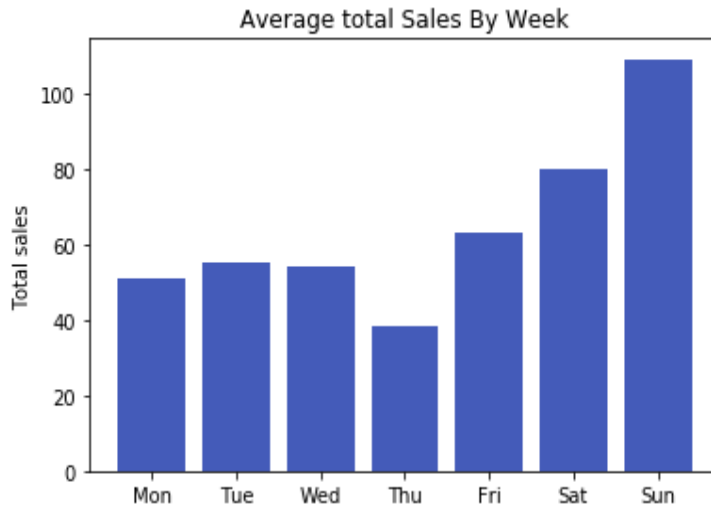


Figure 9 Average Sales based on days of Weeks

Figure 9 illustrate the impact of sales on the basis of days of week. Through this analysis it is concluded that the most significant number of sales were achieve during Sunday throughout the week. While the least number of sales were observed during Thursday throughout the week. It turns out to provide effective policies which can be adapted in order to deal with further more increase the sales.

Another interesting analysis was conducted over the different months of year 2017 and year 2018 against the total number of sales generated in those months shown in Figure 10. Through analysis the fact was find out over the years the greatest number of sales are generated in the month of September. The management team can use this opportunity to create remarkable profit margin if effective polices are implemented precisely. While the least number of sales were noticed in the month of February.

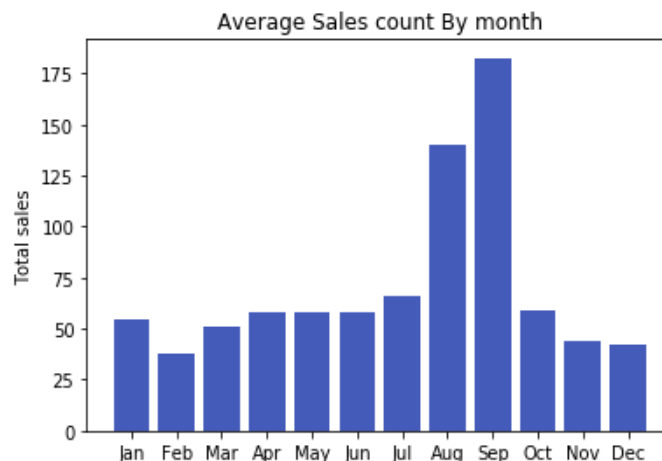


Figure 10 Average Sales based on days of Month

The Analysis conducted over the number of sales over the year shows effective result which tend to provide effective understanding that with a passage of time more and more people are preferring the food of restaurants as shown in Figure 11. Through analysis the fact was concluded that average sales increase 50% as compared to the previous year.

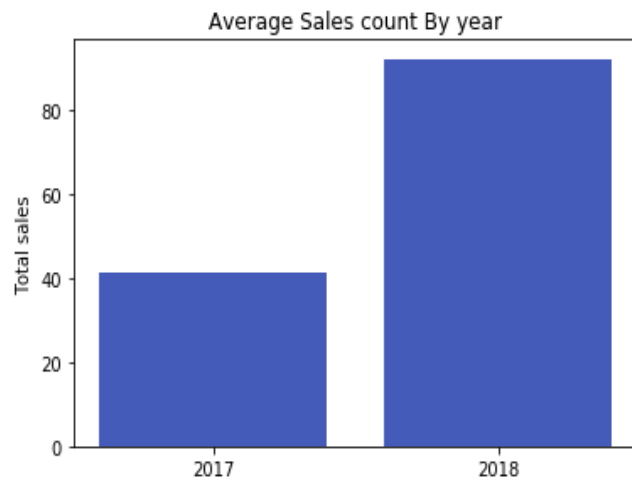


Figure 11 Average Sales based on Yearly Basis

4 Methodology

In order to perform machine learning different methodologies can be adapted to effectively deal with the problem based in real time. In order to achieve the goals of this proposed research and objective behind this research are illustrated in Figure 12 according to(Awad and Khanna, 2015).

4.1 Data Pre-Processing:

The complete exploration of dataset is elaborated in the section 3 of the proposed study. The transforming of data into required information is performed through this part of the research. Pre-process of data is considered as technique of data mining utilized for turning data in raw to a more appropriate format. Removal of missing entries and bulk or null values is performed in this part of the research. while the process of normalization, enrichment and cleaning of data will be performed to trained the model while implementing algorithm based on machine learning. Technique of encoding based on the model of one hot is considered for this research.

4.2 Features Selection:

The selection of those valuable features which can be related to our perspective of research and objective are extracted from the implementation of features selection. The removal of unnecessary features is also performed through this method. The performance was effective improved after the removal of unnecessary features. While the feature of mobile number was utilized as the entity for acquiring unique records. The identification of customer for loyalty is performed through conducting analysis which validate the entity against their respective phone number.

4.3 K-Fold Cross Validation:

The implementation of K-Fold cross validation during machine learning tend to provide effective implementation of data and provide a unique pattern which can easily be understandable. This method tends to split the data into number of K folds. In order to stopping the proposed system in terms of over training only twenty percent of the dataset was utilized in the testing period while

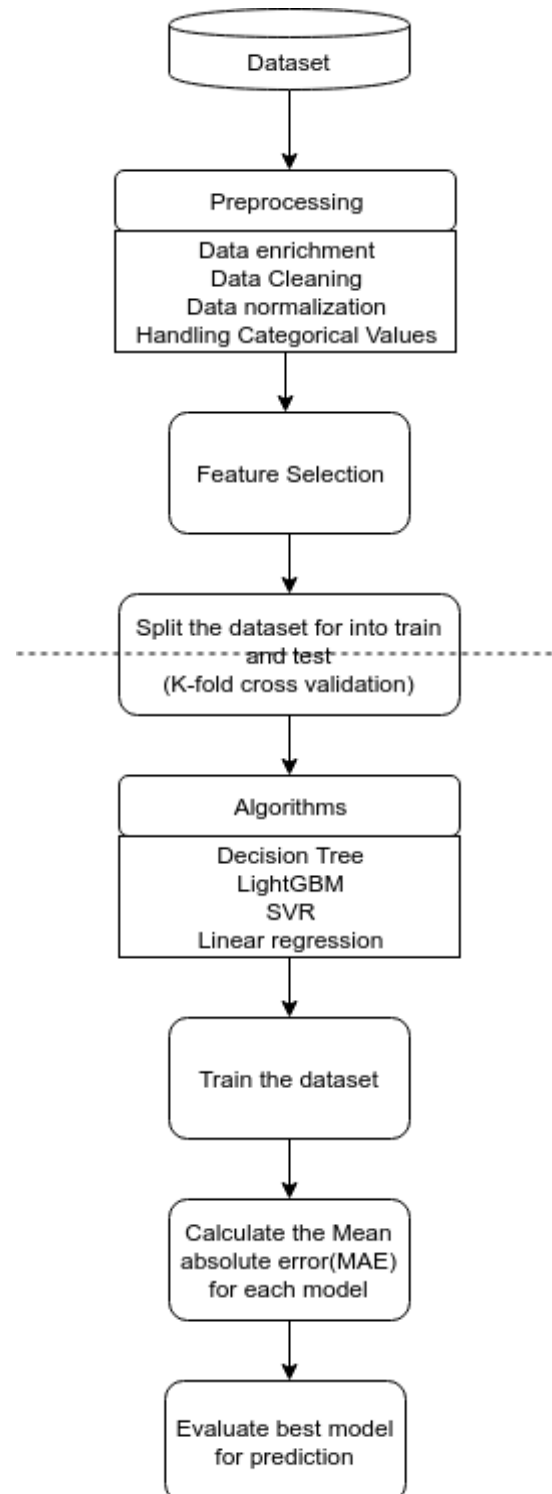


Figure 12 Methodology Adapted in this Study by Awad and Khanna, 2015

eight percent of the dataset was evaluated in training of dataset. Along with effective generalization of the data was achieved through utilization of K-fold cross validation. While the evaluation at each level based on different folds shows effective performance during evaluation.

4.4 Evaluation & Models:

Different algorithms were in this study evaluation period which was based on the following algorithms such as Linear Regression, Support Vector Regression, LightGBM and Decision Tree. After completely the pre-processing of data, selection of different features based on features selection and validation through the means of K-Fold cross validation the complete evaluation of performance along with model training process is considered as the final part of the proposed study. In order to achieve the aims of this proposed research which based on effective prediction for the daily basis sales and total numbers of orders different algorithms were implemented in the proposed study. The validation score turns out to check point which validate the different machine learning till no effective improvement can be observed till straight away two hundred round of time lapse. The evaluation of the proposed study was based on the MAE or Mean Absolute Error for every implemented model of machine learning which were SVR, LR, DT and LGBM. While the performance factor of every algorithm was based on the factor of MAE which will clearly help the researcher to find out the most effective algorithm that can be utilized for effective prediction of sales and orders.

5 Discussion and Results

The result obtained from evaluation of different machine learning algorithm is discussed in this section. The preference of the researcher lies in the meeting the mark up objective of the proposed research which emphasized on the prediction of the revenue which can totally generated on daily basis while which factor can be utilized in order to boost up the number of total sales. While prediction goals also emphasized on the order which can be placed on daily basis. The research implemented different well-known machine learning approaches for the purpose of training and testing of dataset. For evaluation of performance of different learning algorithms based on machines the factor of Mean absolute error is implemented. The programming language of Python was used for the coding scheme for different machine learning algorithms. While the execution of these program was performed on Jupiter notebook. The analysis on total order numbers turned out to be a comparison between different machine learning approaches on the basis of factor of MAE shown in Table 3 which also calculate the performance of different machine learning algorithms.

Algorithm	Mean Absolute Error
Support Vector Regression	34.431
Light Gradient Boosting Machine	15.389
Linear Regression	17.845
Decision Tree	19.120

Table 3 MAE Value bases on Daily Basis total Orders

Algorithm	Mean Absolute Error
Support Vector Regression	10161.840
Light Gradient Boosting Machine	4119.150
Linear Regression	4329.010
Decision Tree	5878.30

Table 4 MAE Value bases on Daily Basis Total Sales

Though evaluation of MAE score in Table 3 of different machine learning algorithms on total number of orders shows the fact that least number of MAE score is secured by algorithm of light Gradient Boosting machine. While the highest score was achieved by SVR during the evaluation of MAE. The main reason behind effective performance of LGBM lies in the fact the algorithm performs to compile data in categorical form which provide advantage as compared to other algorithms.

The analysis on total Sales turned out to be a comparison between different machine learning approaches on the basis of factor of MAE shown in Table 4. Through evaluation the fact was concluded that most satisfying and significantly less encountered were recorded under the implementation of LightGBM technique as compared to other technique. While the least performance in terms high mean absolute error was achieved by Support Vector Regression.

Through different evaluation in section 3 of this research we concluded that cloudy or fair weather boost up the total number of sales and order which can be effectively taken as a vital opportunity to generate extra money in those condition. Foggy weather was turn out to be condition of weather in which customer feels unsecured to order food while introducing different interactive ideas to overcome this issue can be an interactive solution to these issues. Through different evaluation conducted on impact of public holidays on sales and order of restaurant interesting facts were find out. The most significant sales and orders were placed on Sunday which lead us to the fact that restaurant business and management cannot lose this vital opportunity to attract customer in the mean while time. while the least number of customers prefer to use the service of pre-orders. Through evaluation the fact was find out that Zomato is the most used food service used by different users in India. Another interesting analysis was related to extracting the loyal customer were part of this research. The researcher was able to fetch over five hundred top customers on the basis of their loyalty on score generated on the basis of revenue spend by these customers. Different features information related to these loyal customers were part of this evaluation.

6 Conclusion:

The complete overview of the research will be elaborated in this part of this research. This research provided a complete overview of factors which are involved in the effective prediction of sales related to different restaurants. Through this work the effective understanding can be concluded the impact of weather on order and sales of restaurant can be figured out. Different machine learning concepts that can be adapted by different system in order to perform precise prediction is part of the research goals. The vast research provides effective approach to learn modern machine learning algorithms along with its implementation. Different specific method and techniques were imposed on dataset just for effective extraction of data and features in order to utilize those features for the purpose of prediction. While effective prediction was generated after analyzing the existing work based on the total number of sales generated on daily basis along with the total number of orders generated on daily basis. The research highlight majority of customer prefer to pay for expense of restaurant through online services. The greatest number of customers for purpose of online food delivery prefer to use the platform of Zomato. During Public Holiday, especially on Sunday's effective revenue was generated when a very large number of orders and sales are placed. Different algorithms based on machine learning were implemented for the purpose of prediction. Through different

performance evaluation the fact was concluded that the performance achieved by LightGBM surpasses the prediction achieved by any other implemented algorithm on the basis of MAE score. The analysis also identified top first five hundred loyal customer on the base of revenue spend on these platforms. The opportunity will provide a more discount to those customers in order to provide much more effective service to valuable customers. The researcher looks up to other algorithms such as random forest, Naïve bayes and other machine learning and deep learning model for implementing in the same domain of research. Feedback provided by customer is the missing entity in the dataset which can turn out to be effectively improve the performance of any restaurants. In order to achieve effective visualization, the approach of Sentimental analysis should be deployed in the infrastructure for effective analysis.

7 References:

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. Tensorflow: A system for large-scale machine learning, in: 12th $\{\$USENIX\}$ Symposium on Operating Systems Design and Implementation ($\{\$OSDI\}$ 16). pp. 265–283.
- Abrishami, S., Kumar, P., Nienaber, W., 2017. Smart stores: A scalable foot traffic collection and prediction system, in: Industrial Conference on Data Mining. Springer, pp. 107–121.
- Al-Saqqa, S., Al-Naymat, G., Awajan, A., 2018. A large-scale sentiment data classification for online reviews under apache spark. *Procedia Computer Science* 141, 183–189.
- Awad, M., Khanna, R., 2015. Support vector regression, in: *Efficient Learning Machines*. Springer, pp. 67–80.
- Boyapati, S.N., Mummidi, R., 2020. Predicting sales using Machine Learning Techniques.
- Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X., Li, X., 2018. Tada: trend alignment with dual-attention multi-task recurrent neural networks for sales prediction, in: 2018 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 49–58.
- Claypo, N., Jaiyen, S., 2015. Opinion mining for thai restaurant reviews using K-Means clustering and MRF feature selection, in: 2015 7th International Conference on Knowledge and Smart Technology (KST). IEEE, pp. 105–108.
- Francis, H., Kusiak, A., 2017. Prediction of engine demand with a data-driven approach. *Procedia Computer Science* 103, 28–35.
- Holmberg, M., Halldén, P., 2018. Machine Learning for Restaurant Sales Forecast.
- Hossain, F.T., Hossain, M.I., Nawshin, S., 2017. Machine learning based class level prediction of restaurant reviews, in: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, pp. 420–423.
- İşlek, İ., Ögüdücü, Ş.G., 2015. A retail demand forecasting model based on data mining techniques, in: 2015 IEEE 24th International Symposium on Industrial Electronics (ISIE). IEEE, pp. 55–60.
- K.Aishwarya, Aishwarya.N.Rao, Nikita Kumari, Akshit Mishra, Mrs.Rashmi M R, 2020. FOOD DEMAND PREDICTION USING MACHINE LEARNING. *International Research Journal of Engineering and Technology (IRJET)* 07, 3672–3675.
- Kulkarni, Atharva & Bhandari, Divya & Bhoite, Sachin., 2019. Restaurants Rating Prediction using Machine Learning Algorithms. *International Journal of Computer Applications Technology and Research*.
- Lasek, A., Cercone, N., Saunders, J., 2016. Restaurant sales and customer demand forecasting: Literature survey and categorization of methods, in: *Smart City 360°*. Springer, pp. 479–491.
- Liu, X., Ichise, R., 2017. Food sales prediction with meteorological data—a case study of a japanese chain supermarket, in: *International Conference on Data Mining and Big Data*. Springer, pp. 93–104.
- Meulstee, P., Pechenizkiy, M., 2008. Food sales prediction: "If only it knew what we know", in: 2008 IEEE International Conference on Data Mining Workshops. IEEE, pp. 134–143.
- Santhana Lakshmi, V., Bavishna, A., 2020. Analyzing the Restaurant data to Predict the Best Mode of Income. *Sustainable Humanosphere* 16, 270–277.
- Te, Y.-F., Müller, D., Wyder, S., Pramono, D., 2018. PREDICTING THE GROWTH OF RESTAURANTS USING WEB DATA, in: *Economic and Social Development (Book of Proceedings)*, 28th International Scientific Conference on Economic and Social. p. 237.

- Tsoumakas, G., 2019. A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review* 52, 441–447.
- Žliobaite, I., Bakker, J., Pechenizkiy, M., 2009. Towards context aware food sales prediction, in: 2009 IEEE International Conference on Data Mining Workshops. IEEE, pp. 94–99.