Pizza Sales Prediction using Machine Algorithms

Abstract – Pizza sale predicting project aim to utilizing machine learning (ML) techniques for forecasting the sale of different pizza variety and estimation the total income for an given day. In the modern market, with an wide range of pizza option available, it had become challenging for pizza vendor to effective sell they products and for customers to make informed choice. ML offer an promising solution by leveraging large dataset to make accurate predicting and decisions. This project will involve preprocess the pizza sale datasets, including encoding categories variables such as pizza name, size, and categories. The datasets will be splitted into train and test datasets, and an Machine learning model will be Developed using the training date. The trained model will then be used to predict the total income for an given day based on the sale of different pizza variety. The project aim to provide insight into which pizza variety likely to be sell the most and to estimate the total income for an given day, helping. pizza vendor optimize they g sales strategys. The results of this project will demonstrate the effectivenes of using ML techniques in predicting pizza sales and can be used as an basis for furthr research and improvment in sales forecasting models for the pizza industry.

.

I.INTRODUCTION

The significances of predicting pizza sales and estimating total cash cannot be understated in the food industry. With the ever-increasing demands for pizzas and the wide variety of options available, it has become crucial for pizza sellers to accurately foretell sales and optimize their workings. Artificial learning (AL) methods offer promising resolutions in this respect by exploiting historical sales data to strive and build a predictive model utilizing artificial learning methods to generate accurate forecasts and facilitate well-informed decision-making. Our aim in this plan is to exploit preceding sales data to detect underlying patterns and tendencies, thereby permitting us to anticipate fluctuations in pizza sales accurately.. In addition, we can estimate the total income for the day based on these prophecies. The plan will include preprocess the pizza sales dataset, selecting relevant characteristics, and training various AL models such as multiple linear regression and classification algorithms. Through evaluating the efficacy of these models and juxtaposing their preciseness in forecasting pizza sales, we can determine the optimal method for sales projection within the pizza sector. This aims to provide invaluable perspectives to pizza dealers, empowering them to refine sales tactics, streamline inventory management, and thereby elevate overall customer satisfaction. By harnessing the strength of AL methods, we can drive innovation and efficiency in the pizza industry, ensuring its continued success in meeting the evolving demands of clients.

II. LITERATURE REVIEW

Paper - [1] The vehicles' sale predictions, emphasizing on the scarcity of research which focuses on a comprehensive analysis of vehicles' characteristics. The study introduces Fore XGBoost, which is a model designed for precise prediction of vehicle sales! The dataset comprises national car sales data spanning from January 2011 to December 2017, encompassing over 5 million data samples with 33 attributes. ForeXGBoost employs the XGBoost algorithm for prediction, demonstrating superior accuracy compared to linear regression and GBDT models. The model achieves high prediction accuracy with reduced error rates, showcasing its efficacy in precise and efficient vehicle sales forecasting. However, the paper identifies gaps in the incorporation of economic indicators like CPI and GDP, suggesting future research directions to enhance prediction accuracy. The outcome underscores ForeXGBoost's potential for intelligent transportation planning and production forecasting in the automotive industry.

Paper–[2]: The study on Housing Price Prediction via Improved Machine Learning Techniques emphasizes the significance of employing machine learning methods for precise housing price forecasting. It notes that although traditional machine learning techniques are widely used, they often overlook individual model performance and neglect complex models. The research aims to investigate how various features impact prediction methods by applying both traditional and advanced machine learning approaches. The dataset utilized in the study is the "Housing Price in Beijing" dataset, comprising over 300,000 data points with 26 variables representing housing prices traded between 2009 and 2018. These dataset features were used to predict the average price per square meter of each house. Data preprocessing involved handling missing data, conducting feature engineering processes, and detecting outliers using the Inter-Quartile Range (IQR) method. The study compares different machine learning models for housing price prediction, such as Random Forest, XGBoost, LightGBM, Hybrid Regression, and Stacked Generalization Regression.The Stacked Generalization approach serves as a machine learning ensemble technique to enhance predicted values. The evaluation metric employed is the Root Mean Squared Logarithmic Error (RMSLE), with the Stacked Generalization method achieving the lowest RMSLE of 0.16350 on the test set. One potential gap identified in the paper is the insufficient discussion on specific features or attributes that significantly influence housing price prediction. Although the study mentions feature engineering processes, providing more insights into the importance of certain features and their impact on predictive models would be beneficial. The research date findings illustrates that Slopped Generalization Regress method excel past the Hybrid Regression method in generalization. The study concludes with discourse on the performance of various models, emphasizing their strengths and weaknesses. It indicates that although all methods produce satisfactory results, individual modeling has its own pros and con downsides. Generally speaking, this study provides invaluable insights over the application of mechanism knowledge techniques for prediction of housing prices and hints on promising pathways for additional study in this area.

Paper – [3]:"Price Cop–Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform" provides an in-depth look at price monitoring and prediction within the e-commerce industry. It depends on a dataset of product prices sourced from e-commerce platforms to train and evaluate the models. The dataset's sufficiency is crucial to ensure accuracies of price prediction and monitoring in Price Cop. However, it limits concerning historical data impacts the training process of the LSVM-ABC model. The study uses Linear Regression (LR) to anticipate product prices on the e-commerce platform, juxtaposing it with the Least Squares Support Vector Machine (LSSVM) and Artificial Bee Colony (ABC) methods. The LSVM-ABC model shows higher accuracy (84%) in price prediction than LR (62%), demonstrating its superior performance. Additionally, the LSVM-ABC model fine-tunes using ABC to enhance results in each iteration. Accurately predicting prices essential for users to make well-informed purchase decisions, even though the study acknowledges the dataset's limitations hampering the LSVM-ABC model's full potential. It proposes the need for a more extensive dataset with daily updates to better prediction accuracy. Also, the study recognizes the gap between predictive and prescriptive analytics as a potential area for future examination. PriceCop introduces as a web application crafted to aid users in monitoring and predicting products prices on e-commerce platforms, offering characteristics for price predictions to help users analyzing price trends and planning purchases. The study emphasizes the importance of user-friendly systems for monitoring prices counter misconception pricing tactics utilized by online retailers. PriceCop aims to empower users to make informed decisions and evade overpaying for products to deceitful pricing strategies. In conclusion, the paper offers valuable insights into the development of PriceCop and the use of predictive models like LR and LSVM-ABC for price monitoring in e-commerce. It emphasizes significant of accurate price prediction and acknowledges dataset limitations and advocates for further research to improve predictive analytics in e-commerce sector.

Paper – [4]: The article on Building Value Guessing Design Utilizing Devices Learning plunges into the spread of Tools Learning and its applications, with an explicit emphasis on building value guessing as a conspicuous domain of interest. Past examination, embodied by academia like Patel and Upadhyay, has plunged into pruning strategies and precision valuation utilizing the TWEEZERS instrument, investigating datasets like glass and diabetic. The use of selection shrub tactics like ID3 and TDIDT for making arrangement norms via decision shrubs is underscored. A real-opportunity dataset is curated through an analysis of the locale of Tadepalligudem in West Godavari Sector, Andhra Pradesh, India. Critical aspects of this dataset encompass the figure of sleeping chambers, building age, transit facilities, closeness to schoolhouses, and entrance to shopping comforts. The suggested strategy applies decision shrub categorization, decision shrub recession, and various horizontal recession for building value guessing. The decision shrub filer is employed to prophesy building availability based on user prerequisites, whilst recession technologies are used for guessing building rates. The accuracy of the decision shrub model is assessed utilizing 20% of the dataset as examination data, with activity appraised through metrics like Nasty Quadratic Mistake (NME), Nasty Absolute Mistake (NE), and flip nasty quadrupled slip (RNSE). Although the manuscript does not openly pinpoint any distinct examination gaps, it gestures at possible future improvements by enriching datasets with additional facets and exploring enhanced devices learning methodologies for building value guessing. The developed method facilitates the guessing of building availability and rates based on user restraints, thereby supporting users in making well-taught resolutions regarding domestic acquisitions. The examination concludes that numerous horizontal recession outperforms decision shrub recession in building value guessing precision, suggesting that future examination could engage dataset refinement and the assumption of enhanced devices learning methods.

Paper – [5]: Analysing Of B2B Sales Forecasting For Telecommunication Facility Using Machine Learning Practices. The research delve within dissecting B2B sales information to support telecom firms in effective managing their sales troops, wares, and money distributions. Multiple machine learning methods like decision trees, haphazard forests, and gradient enhanced trees are leveraged to hone sales prognostication precision. The exploration uses B2B sales statistics related to telecommunications services for scrutiny. A variety of machine learning models, taking in decision trees, random forests, and gradient boosted trees, are put into service to foretell forthcoming sales inclinations. The competency of these models is evaluated through metrics like Mean Squared Error (MES) and Mean Absolute Percentage Error (MAPE). Among these, the gradient boosted tree model showcases top-notch performance, reaching an MES of 24,743,000,000.00 and a MAPE of 0.18, indicating remarkable accuracy in foretelling approaching B2B sales. The paper could possibly profit from further exploration of alternative machine learning algorithms or ensemble techniques to refine and compare predictive exactness. Incorporating a wider array of aspects or extraneous variables that might impact sales inclinations could generate a more extensive analysis. The study underscores the pivotal function of machine learning methods in refining sales forecast precision for telecommunications services. The gradient boosted tree model arises as the best selection for accurately prophesying B2B sales, emphasizing the significance of deploying advanced predictive models to enhance sales forecasting and decision-making in the telecommunications industry.

III. PROBLEM DEFINATION

An aim be to predicting the most selling pie and calculating overall revenue per day with pizza sales datasheet!!!

DATASET

In this here project, our dataset was found on Kaggle, which has 48,620 rows of data spread out in 12 row-ups showing different stuff. The main aim is to make a guess-testing model using these 12 factors to predict the value of pizza!

DATA PRE-PROCESSING

Data preprocessing for pizza sales prediction involves several steps to ensure the dataset is ready for machine learning models. These steps include handling missing values by filling them with appropriate strategies, encoding categorical variables into numerical format, and Data prepossessions for pizza sail predicament includes multiple strides to ensure the datasets are ready for machinery learning models. These strides include handling miss values by filling them with appropriate strategics, coding categorial variables into numerical formats, and performing characteristics engineering to create newly related characteristics. Furthermore, normalizing or standardizing numerical characteristics, extracting data functions from order date column, and handling out layers are significant. The datasets should be divided into trainings, validations, and tests sets, and feature selections can be performed to choose most related features. Lastly, if the datasets are unbalanced, techniques like over sampling or under sampling can be applying.

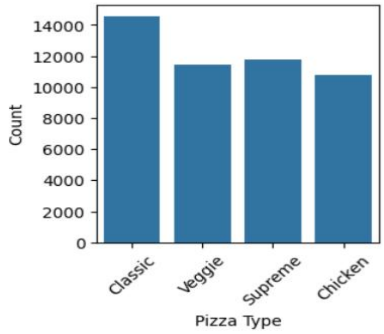

Fig.1. pizza_category attribute

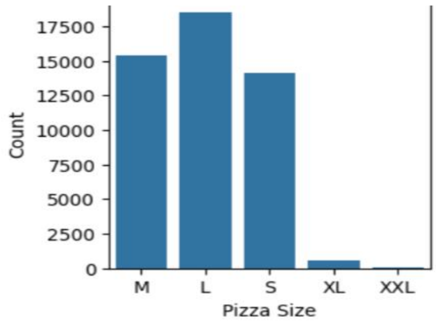

Fig.2. Pizza Size

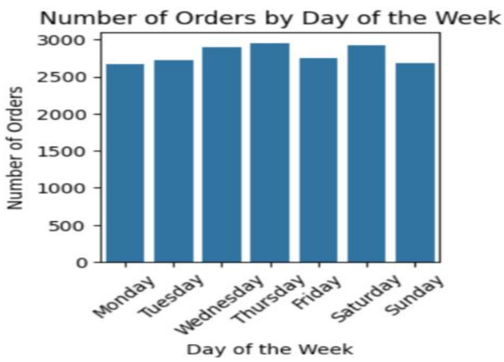

Fig.3. Order_date

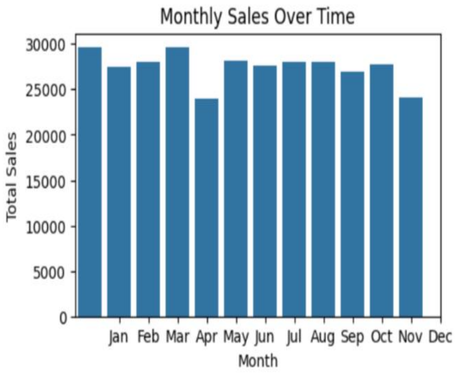

Fig.4. Monthly Sales

Figure 1 to 4 show the outcomes of attribute preprocessing done on the dataset. Figure one illustrate the selected attributes for the pizza Group, while Figure 2 portray those for sizes of Pizzas. Figure 3 show the attributes for Weekly Sales, and Figure 4 represents the yearly Monthly Sales data.

C. ALGORITHMS

We implement a variation of Machinery Learning (ML) models, including various Regression models, to address our prediction duties. This implied preparing data for training and testing the ML models.

Decision Tree: Decision Tree are supervised learning algorithm utilized for classification and regression duties. In predictable pizza sales, Decisions Tree can be trained to examine multiple factors (e.g., pizza size, category, ingredients) and forecast the sales quantities. It functions by constructing tree-like model where each internal node signifies a "decision" based on an

attribute, each branch signifies the result of decisions, and each leaf nodes denotes the ultimate predictions (sales quantities). Decision Trees are known for their interpretability and abilities to handle both numeric and categorial data, rendering them suitable for annual pizza sales data.

Random Forest: Random Forest are an ensemble learning technique that enhances prediction accuracies by employing multiple Decisions Trees. In the pizza sales prediction project, Random Forests can be applied to construct forests of Decisions Trees, with each trees trained on a random subset of the data. By combining the predictions from numerous trees, Random Forest can offer more precise predictions compared to a single Decision Tree. This approach is adept at handling extensive datasets and is resilient against overfitting, rendering it an appropriate option for forecasting pizza sales.

Linear Regression: Linear Regression are a commonly used and straightforward regression algorithms that establishes a relationships between a dependent variables (such as pizza sales quantities) and one or more independent variable (features like pizza size, price, etc.) by fitting a linear equation to the available data. In the realm of predictable pizza sales, Linear Regression can be employed to anticipate sales quantities based on the provided features. It presuppose a linear connection between the features and the target variable, thus ensuring ease of interpretations and applicability to datasets exhibiting approximately linear relationships.

Gradient Boosting Machines (GBM): Gradient Boosting Machines (GBM) is a ensemble learning method that constructs a robust predictive models by amalgamating the forecasts of numerous weak models, typically decision trees. GBM iteratively constructs trees, with each trees rectify the errors of its predecessor, culminating in a more precise and resilient model. In the realm of predicting pizza sales, GBM can be utilized to forecast sales quantities based on diverse features. Renowned for its high accuracy and adeptness in handling intricate data relationships, GBM proves beneficial in enhancing predictions performance compared to individual decision trees or linear models.

K-Nearest Neighbours (KNN): KNN, or k-Nearest Neighbours, represents a straightforward, instance-based learning techniques applicable to both classifications and regression objectives. In the domain of predictable pizza sales, KNN offers the capabilities to anticipate sales quantities by assessing the similarity among pizza orders. The algorithms operates by identifying the k nearest neighbours (pizza orders) to a specific orders within the training dataset and averaging their sales quantities to forecast the sales volumes of the given order. KNN's simplicity and ease of implementations render it a appropriate selections for small to medium-sized pizza sales datasets.

Neural Network: Neural networks represents a category of machinery learning models inspired by the structure and functionalities of the human brain. When applied to forecasting pizza sales, a neural networks can discern intricate patterns within the data and generate predictions based on these patterns. Comprising layers of interconnected nodes (neurons), a neural networks processes inputs data and transmits the output to subsequent layers. The initial layer (input layer) receives the input features (e.g., pizza size, category, ingredients), while intermediate layers (hidden layers) analyse the input, and the final layer (output layer) yields the predictions (sales quantities). Notably, neural networks possesses the capabilities to learn from data autonomously, enabling them to tackles complex data relationships effectively. They excels in scenarios where the association between inputs and outputs is nonlinear and intricate, rendering them well-suited for predictable pizza sales.

XG Boost: XG Boost, a optimized gradient boosting algorithms, proves highly effective for fore casting pizza sales. It's adeptness in managing intricate data relationships makes it an excellent choices for capturing the subtleties of pizza sales patterns influenced by diverse factors like pizza size, category, and ingredients. The key advantages of XG Boost lies in its sequential addition of decisions trees, with each trees refining the errors of its predecessors, culminating in a robust model capable of accurately predicting daily pizza sales. Furthermore, its efficiency and rapid processing makes it well-suited for handling large datasets, empowering businesses to make informed decisions regarding inventory managements and sales strategies. Through leveraging XG Boost, businesses can optimized their operations, trim expenses, and enhance customer satisfactions by ensuring timely availability of the right pizzas based on predicted demands.

Ridge regression: Ridge regression is a regularization techniques used to prevent overfitting in linear regression models. It's suitable for the pizza sales prediction project as it can handles multicollinearity (when predictor variables are correlated) and preventing the model from being overly influenced by any single variable. This is important when dealing with attributes like pizza size, category, and ingredients, which might be correlated. Ridge regression adds a penalty term to the linear regression equation, which helps to stabilize the model and improve its generalizations performance. Overall, Ridge regression can enhances the accuracies and robustness of the prediction models for pizza sales.

IV. BUILDING THE MODEL

In them segment, wee perforated essential data preprocessing steps two prepare the dataset fur machine learning modeling in pizza seals predicament. Wee converts some categorical elements into numeric values, witch is crucible fur training machine learning models!!!!The dataset contains information on pizza sales, including pizza_id, order_id, pizza_name_id, quantity, order_date, order_time, unit_price, total_price, pizza_size, pizza_category, pizza_ingredients, and pizza_name.

Starting by taking out any columns that aren't helping with predicting tasks. Then, we pull out features like what day, month, or year it is from the order_ date to catch any trends in pizza sales that change with the season or time. We might also make some new features using what we know about the field, like how many ingredients are in a pizza or the mean price for each ingrediences

Moreover, we address categorical variables by transforming them into numerical values through methods such as one-hot encoding or label encoding. This ensures that machine learning models can properly interpret these variables. Finally, we partition the dataset into training and testing subsets to assess the effectiveness of our models.

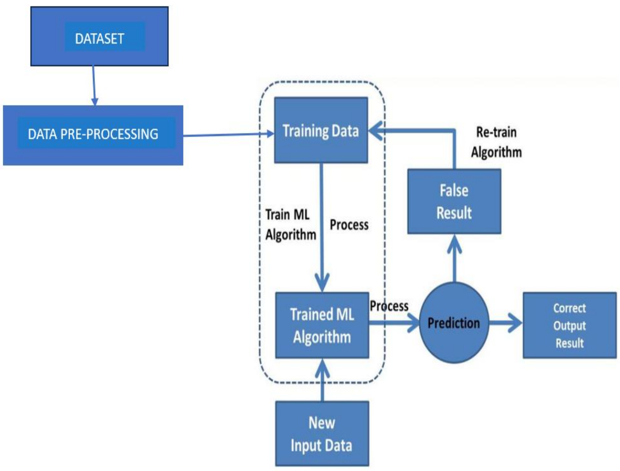

Fig:5 Model Building

PREDICTIONS

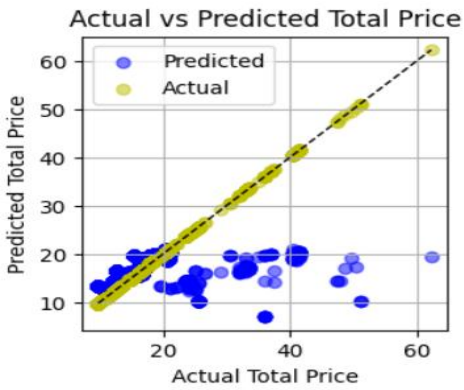

Fig.6. linear regression

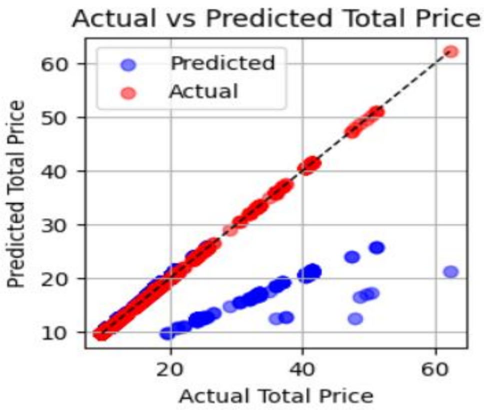

Fig.7. Decision Tree

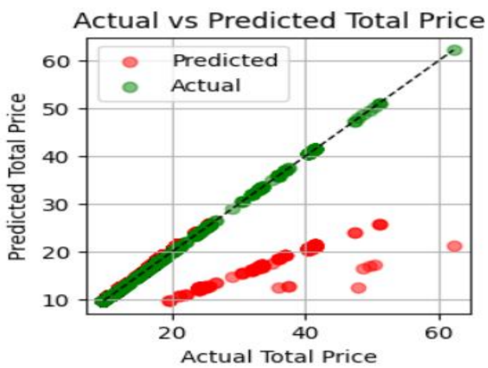

Fig.8.RandomForest

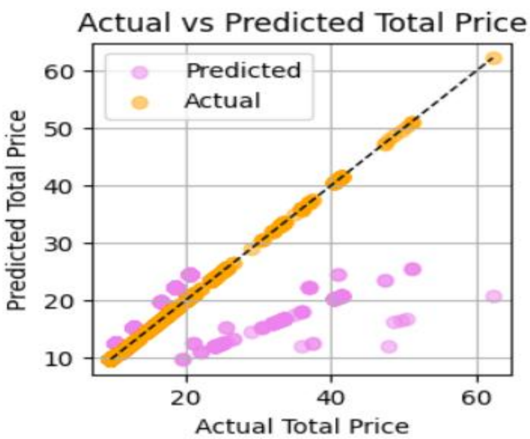

Fig9 .KNN Regression

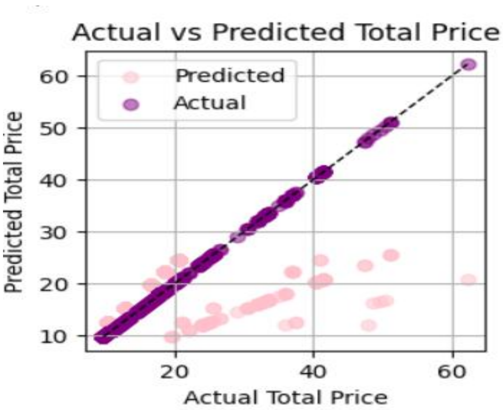

Fig10.Nueral Network

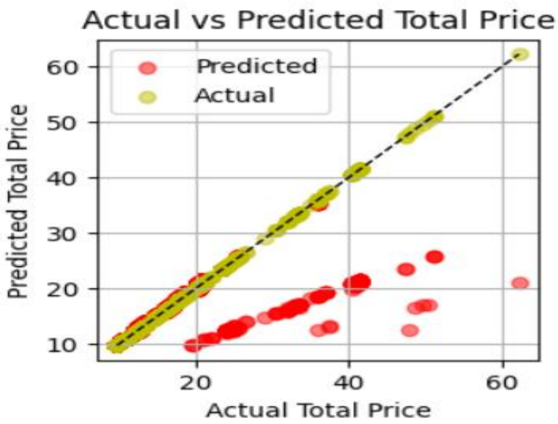

Fig 11.GradientBoostingRegressor

Fig12. XG BOOST

Fig 13. Ridge Regression

The aforementioned figures illustrate the comparison between the actual values and predicted values, indicating the model accuracy across the various implemented models. We assess the accuracy of the models by predicting values from the test data. The predictions generated by different models were evaluated to gauge their performance.

Model	R-squared	Mean Squared Error(MSE)	Root Mean Squared Error(RMSE)	Mean Absolute Error (MAE)	Mean Absolute Error Percentage (MAPE)	Mean Squared Error Percentage
Linear Regression	0.39320	12.1435	3.48471	1.64263	9.79%	72.37%
Decision Trees	0.69503	6.10312	2.47044	0.64284	3.83%	36.37%
Random Forests	0.69503	6.10321	2.47046	0.64365	3.84%	36.37%
KNN Regression	0.65665	6.87129	2.62131	0.56444	3.36%	40.95%
Neural Networks	0.65665	6.87129	2.62131	0.56444	3.36%	40.95%
Ridge Regression	0.39320	12.1435	3.48475	1.64268	9.79%	72.37%
XG Boost	0.69503	6.10312	2.47044	0.64285	3.83%	36.37%
GBM	0.69338	6.13610	2.47711	0.67347	4.01%	36.57%

Fig.9 Results of the models

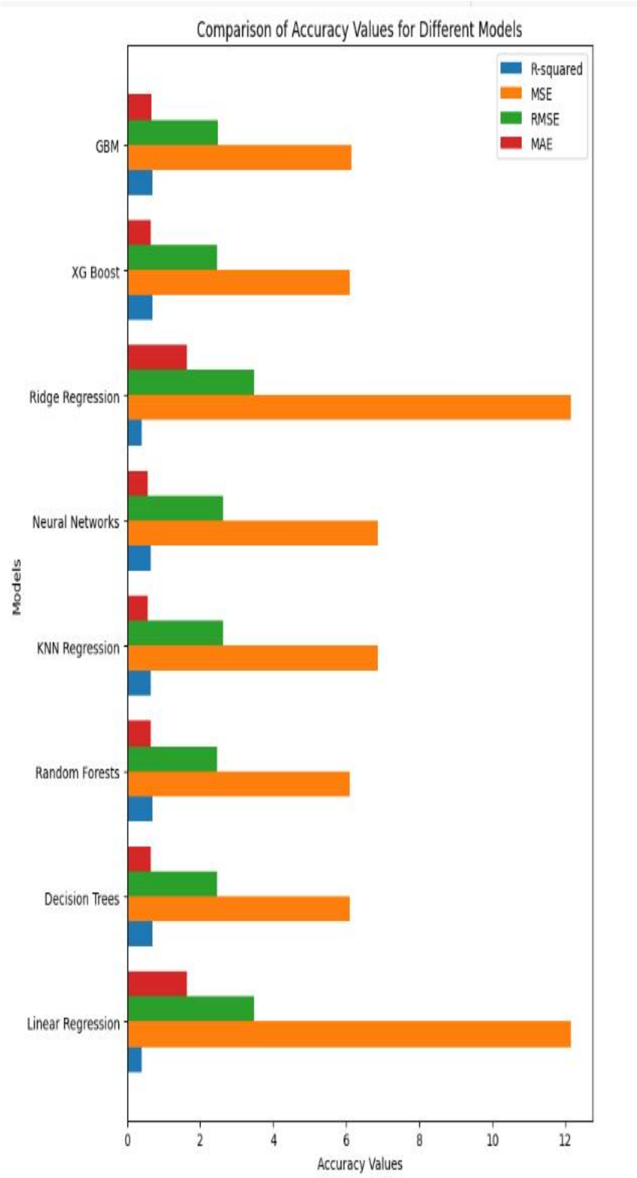

Fig 10 .comparison of Accuracy

The provided figures depict the outcomes derived from the pizza sales predictor platform. Our analysis involved employing algorithms such as linear regression, random forest, decision tree, KNN, Neural Network, Ridge Regression, XG Boost, and GBM. As we can see from the above table, the Decision Trees and Random Forests models seem to perform the best, with the highest R-squared values (0.69503), indicating a better fit to the data compared to other models. They also have relatively low Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) values, suggesting that they make more accurate predictions compared to other models. XG Boost also performs well, with similar metrics to Decision Trees and Random Forests. Therefore, Decision Trees, Random Forests, and XG Boost are the best-suited models for the pizza sales prediction project.

V. CONCLUSION

The pizza sales predictor project utilized various machine learning algorithms to forecast pizza sales based on several factors such as pizza type, quantity, order date, and order time. Through the analysis, it was found that the Random Forest and Decision Trees models performed the best, with the highest R-squared values and lowest error metrics compared to other models like Linear Regression, KNN, and Neural Networks. The models Predicted, Predicted Price is 16.821 Most Sold Pizza is The Classic Deluxe Pizza. Overall, the project successfully showcased the applicability of machine learning in predicting sales patterns, which can be valuable for businesses in the food industry for inventory management and resource allocation.

REFERENCES

[1] Aminah Md Yusof and Syuhaida Ismail, Multiple Regressions in Analyzing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.

[2] Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to overfitting regression-type models. Psychosomatic Medicine, 66(3), 411-421.

[3] Atharva shoogles, Priyanka chair, Akshata gaud, Jinal Jain.House Price Forecasting using Data Mining Techniques International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 12, December 2017

[4] Model, Azmi Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin, Comparative Study On Estimate House Price Using Statistical And Neural Network, International journal of scientific and technology, research volume 3, ISSUE 12, December 2014,Page(s):126-131.

[5] Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, Wayan Firdaus Mahmudy,Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study: Malang, East Java, Indonesia. (ITASCA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017, Page(s):323-326..

[6] S. Jyothirmayee, V. D. Kumar, C. S. Rao, and R. S. Shankar, "Predicting Stock Exchange using Supervised Learning Algorithms," no. 1, pp. 4081-4090, DOI: 10.35940/invitee.A4144.119119.

[7] H. Snyder, "Literature review as a research methodology: An overview and guidelines," J. Bus. Res., vol. 104, no. 2019, pp. 333339, DOI: 10.1016/j.jbusres.2019.07.039.

[8] N. S. Selamat, "An Overview of Big Data Usage in Disaster Management," vol. 11, no. 1, pp. 35-40.

R. P. Schumaker and H. Chen, “Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System,” no. 2006, pp. 1 -29.

.