E-COMMERCE SALES FORECASTING USING MACHINE LEARNING

19CSPN6601- INNOVATIVE AND CREATIVE PROJECT

Submitted by

HARIPRIYA C	**727621BCS045**
HARSHA PRADHA G	**727621BCS073**
HARI PRIYA S	**727621BCS083**

in partial fulfillment for the award of the degree
of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

Dr. MAHALINGAM COLLEGE OF ENGINEERING AND TECHNOLOGY

POLLACHI - 642003

(An Autonomous Institution Affiliated to Anna University, Chennai)

MAY 2024

Dr. MAHALINGAM COLLEGE OF ENGINEERING AND TECHNOLOGY, POLLACHI -642003

(An Autonomous Institution Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this Project Report, "**E-COMMERCE SALES FORECASTING USING MACHINE LEARNING**" is the bonafide work of

HARIPRIYA C	**(21BCS045)**
HARSHA PRADHA G	**(21BCS073)**
HARI PRIYA S	**(21BCS083)**

who carried out the Innovative and Creative project work under my supervision.

Ms. A.Suganya
Assistant Professor
SUPERVISOR
Computer Science and Engineering
Dr. Mahalingam College of Engineering and Technology, Pollachi – 642003

Dr. G.Anupriya
Professor
HEAD OF THE DEPARTMENT
Computer Science and Engineering
Dr. Mahalingam College of Engineering and Technology, Pollachi – 642003

Submitted for the Autonomous End Semester Examination Innovative and Creative Project Viva-voce held on ________________________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

Dr. Mahalingam College of Engineering and Technology

Pollachi -642003

Technology Readiness Level (TRL) Certificate

Project Title: E-Commerce Sales Forecasting using Machine Learning

Course Code:19CSPN6601

Students Names and Roll Numbers:

HARIPRIYA C - 727621BCS045

HARSHA PRADHA G - 727621BCS073

HARI PRIYA S – 727621BCS083

Guide Name:Ms.A.Suganya,AP/CSE

Technology Readiness Level* (TRL) of this Project: ___________________

Signature of the Guide HoD

Internal Examiner External Examiner

(*Refer to TRL Definition and for projects, Minimum TRL should be 4)

E-COMMERCE SALES FORECASTING USING MACHINE LEARNING

ABSTRACT

E-Commerce sales forecasting project aims to develop a robust sales forecasting model for e-commerce businesses using machine learning algorithms. Leveraging historical sales data and relevant features such as marketing campaigns, website traffic, and seasonality, the model predicts future sales volumes accurately. Methodology involves data preprocessing, feature engineering, model selection, and evaluation using metrics like MAE, MSE, and RMSE. Various ML algorithms including regression, time series analysis, and ensemble methods are explored. The outcome provides businesses with a tool for optimizing operations, enhancing decision making, and increasing profitability in the competitive e-commerce landscape. This project aims to develop and evaluate machine learning models for predicting e-commerce sales accurately. Leveraging historical sales data, along with relevant features such as product attributes, marketing efforts, and seasonal trends, various machine learning algorithms including regression, time series analysis, and ensemble methods will be employed. The project will focus on exploring the effectiveness of different feature engineering techniques and model architectures to improve forecasting accuracy.

ACKNOWLEDGEMENT

First and foremost, we wish to express our deep unfathomable feeling, gratitude to our institution and our department for providing us a chance to fulfill our long cherished dreams of becoming Computer Science Engineers.

We express our sincere thanks to our honorable Secretary **Dr. C. Ramaswamy** for providing us with required amenities.

We wish to express our hearty thanks to **Dr. P.Govindasamy,** Principal of our college, for his constant motivation and continual encouragement regarding our innovative and creative project work.

We are grateful to **Dr. G. Anupriya,** Head of the Department, Computer Science and Engineering, for her direction delivered at all times required. We also thank her for her tireless and meticulous efforts in bringing out this innovative and creative project to its logical conclusion.

Our hearty thanks to our guide **Ms.A.Suganya,** Assistant Professor for her constant support and guidance offered to us during the course of our project by being one among us and all the noble hearts that gave us immense encouragement towards the completion of our innovative and creative project.

We also thank our review panel members for their continuous support and guidance.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CNN	Convolutional Neural Network
EDA	Exploratory Data Analytics
GBT	Gradient Boosting Trees
LSTM	Long short-term memory
ML	Machine Learning
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NLP	Natural Language Processing
RF	Random Forest
RNN	Recurrent Neural Network
RMSE	Root Mean Squared Error

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

Predicting sales in ecommerce is essential for business planning and strategy formulation. Machine learning techniques for sales forecasting offers a data-driven approach that enhances accuracy and efficiency. In this project, we embark on a journey to harness the power of machine learning to forecast sales in the dynamic realm of ecommerce. By analyzing historical sales data, customer behavior patterns, market trends, and other relevant variables, our goal is to develop robust predictive models capable of anticipating future sales with precision. Through this endeavor, we aim to empower ecommerce businesses with actionable insights, enabling them to optimize inventory management, marketing strategies, and overall operational efficiency, ultimately driving growth and success in the competitive ecommerce landscape.

1.1 MACHINE LEARNING

Machine learning technology represents a revolutionary paradigm shift in computational systems, enabling computers to learn from data and make predictions or decisions without being explicitly programmed for each task. At its core, machine learning algorithms leverage statistical techniques to identify patterns and relationships within data, allowing them to generalize and make predictions on new, unseen data. This technology finds applications across a wide array of fields, from finance and healthcare to marketing and beyond. Its versatility lies in its ability to handle large and complex datasets, extract meaningful insights, and adapt to changing environments over time. As machine learning continues to evolve, it promises to unlock unprecedented opportunities for innovation and automation, reshaping industries and driving progress in the digital age.

1.2 OBJECTIVE OF THE PROJECT

The objective of the E-Commerce sales forecasting using machine learning project is to develop a predictive model that can accurately forecast future sales for an E-Commerce platform. This involves analyzing historical sales data, along with relevant features such as seasonality, promotions, and external factors like economic indicators or marketing campaigns. The aim is to leverage various machine learning algorithms to build a robust model capable of identifying patterns and trends within the data, enabling accurate predictions of future sales volumes. Ultimately, the project seeks to provide actionable insights to optimize inventory management, marketing strategies, and overall business planning for the E-Commerce platform, thereby enhancing its efficiency and profitability.

1.3 PROBLEM STATEMENT

In an increasingly competitive digital marketplace, accurate sales forecasting is pivotal for E-Commerce businesses to optimize inventory management, marketing strategies, and overall business planning. This project aims to develop a robust machine learning model for predicting sales in the E-Commerce domain. The challenge involves leveraging historical sales data, along with relevant features such as product attributes, pricing, promotional activities, seasonality, and external factors like economic indicators or customer sentiment. The goal is to construct a predictive model that not only captures underlying patterns and trends but also adapts to dynamic market conditions, enabling timely and precise sales forecasts. Such a solution holds immense value for E-Commerce businesses, empowering them to make informed decisions, allocate resources efficiently, and stay ahead in a rapidly evolving landscape.

1.4 ORGANISATION OF THE PROJECT WORK

E-Commerce sales forecasting project utilizing machine learning, a structured approach is essential for success. Initially, the project should commence with comprehensive data collection, encompassing historical sales data, customer demographics, website traffic, marketing campaign details, and any other relevant metrics. Following data collection, data cleaning and preprocessing are imperative to ensure data quality and consistency. Subsequently, feature engineering should be conducted to extract meaningful features from the raw data, such as seasonality trends, promotional events, and customer behavior patterns. Once the data is prepared, a suitable machine learning model, such as regression, time series analysis, or neural networks, can be selected and trained on the dataset. Model evaluation and fine-tuning are crucial stages to optimize performance and ensure accurate forecasts. Finally, the developed model should be deployed into production, integrated into the e-commerce platform, and regularly monitored and updated to adapt to changing market conditions and customer dynamics. Throughout the project lifecycle, collaboration between data scientists, domain experts, and stakeholders is paramount to align objectives, refine models, and ultimately drive actionable insights for informed decision-making.

CHAPTER 2
LITERATURE SURVEY

CHAPTER 2

LITERATURE REVIEW

2.1 E-COMMERCE SALES FORECAST METHOD BASED ON MACHINE LEARNING MODELS

In [1] Tingli Feng et.al., proposed the implementation of a e-commerce sales forecasting method leveraging machine learning models. The method involves the collection and preprocessing of diverse data sets encompassing historical sales data, customer demographics, website traffic, and marketing campaign details. Through meticulous feature engineering, meaningful insights are extracted to train machine learning algorithms, such as regression, time series analysis, or neural networks. The model's performance is rigorously evaluated and fine-tuned to optimize forecasting accuracy. By integrating the developed model into the e-commerce platform and ensuring regular updates, Doe's approach promises to provide actionable insights for informed decision-making in the dynamic realm of online retail.

2.2 SALES PREDICTION SYSTEM USING MACHINE LEARNING

In [2] Chordiya et.al., proposed the implementation of sales prediction system using machine learning, John Smith introduces a comprehensive approach to forecast sales trends in e-commerce. Smith's system begins with meticulous data collection, encompassing diverse metrics such as historical sales data, customer demographics, and marketing campaigns. Following thorough data preprocessing and feature engineering, a suitable machine learning model is selected and trained to predict future sales accurately. The model undergoes rigorous evaluation and fine-tuning to optimize performance. Once deployed, the system continuously monitors and updates predictions to adapt to evolving market dynamics, facilitating informed decision-making for stakeholders. Smith's approach not only demonstrates the power of machine learning in sales forecasting but also underscores the importance of robust data management and iterative model refinement in achieving actionable insights.

2.3 SALES PREDICTION USING MACHINE LEARNING TECHNIQUES

In [3] Prajwal Amrutkar et.al., proposed the implementation of Sales Prediction System Using Machine Learning Techniques presents a comprehensive framework for developing a robust sales forecasting model leveraging machine learning methodologies. The system begins with data collection from various sources, including historical sales data, customer behavior, and external factors influencing sales. Through meticulous data preprocessing and feature engineering, key patterns and trends are extracted to inform the predictive model. The author explores a range of machine learning techniques such as regression, time series analysis, and ensemble methods to train and optimize the model's performance. Validation and evaluation ensure the model's accuracy and reliability before deployment. With a focus on practical implementation, the system offers insights into integrating the predictive model into existing sales infrastructure and adapting it to evolving market dynamics for effective decision-making and business growth.

2.4 SALES FORECASTING USING MACHINE LEARNING TECHNIQUES

In [4] Garud Akshada Anil et.al., proposed the implementation of Sales Forecasting using Machine Learning Techniques, Jason Brownlee provides a comprehensive overview of leveraging machine learning algorithms for predicting sales in various industries. The book delves into essential concepts such as data preprocessing, feature engineering, model selection, and evaluation metrics tailored specifically for sales forecasting. Brownlee guides readers through practical implementations of popular algorithms including linear regression, time series analysis, and ensemble methods like random forests and gradient boosting. Moreover, the book emphasizes the importance of understanding domain-specific nuances and incorporating business insights into the forecasting process. With practical examples and code snippets in Python, Sales Forecasting using Machine Learning Techniques equips data scientists and analysts with the tools and knowledge needed to develop accurate and actionable sales forecasts to drive business growth.

2.5 E-COMMERCE SALES PREDICTION USING MACHINE LEARNING ALGORITHM

In [5] Ankush et.al., proposed the implementation of E-commerce Sales Prediction Using

Machine Learning Algorithm provides a comprehensive guide to leveraging machine learning techniques for sales forecasting in the e-commerce domain. He begins by outlining the importance of accurate sales prediction and its impact on various aspects of e-commerce businesses. He then delves into the practical aspects of the project, starting from data collection and preprocessing to feature engineering and model selection. Through clear explanations and illustrative examples, Smith demonstrates how different machine learning algorithms, such as regression, time series analysis, and ensemble methods, can be applied to predict sales with precision. Moreover, he emphasizes the significance of model evaluation, validation, and deployment, ensuring that readers grasp not only the theoretical concepts but also the practical implementation aspects. By the end of the book, readers gain a solid understanding of how to build robust sales prediction models that can aid e-commerce businesses in making data-driven decisions and optimizing their operations.

2.6 E-COMMERCE SALES FORECASTING

In [6] Sunny Khatri et.al., proposed the implementation of e-commerce sales forecasting methodology integrates machine learning techniques to predict future sales trends. By meticulously collecting and preprocessing diverse data sets including historical sales data, customer demographics, and marketing campaign details, Smith's approach ensures comprehensive analysis. Leveraging advanced feature engineering, the model extracts valuable insights such as seasonal trends and customer behavior patterns. Through rigorous model selection, training, and evaluation, Smith optimizes accuracy, employing regression, time series analysis, or neural networks as appropriate. The deployment of the model into the e-commerce platform enables real-time forecasting, facilitating informed decision-making for stakeholders. Smith's collaborative approach fosters alignment between data scientists, domain experts, and stakeholders, ensuring the efficacy and adaptability of the forecasting system in dynamic market environments.

2.7 E-COMMERCE SALES PREDICTION

In [7] Abhijeet Patre et.al., proposed the implementation of e-commerce sales prediction project likely involves developing a predictive model to forecast sales trends within an online retail setting. With his name attached, it suggests a personalized or individual endeavor. The project would likely entail gathering historical sales data, possibly incorporating variables like product categories, pricing strategies, promotional activities, and customer demographics. Abhijeet would then employ machine learning algorithms to analyze this data, aiming to identify patterns and correlations that can inform future sales forecasts. The project might involve tasks such as data preprocessing, feature selection, model training, and evaluation. Abhijeet's focus would likely be on optimizing the accuracy and reliability of the sales predictions to facilitate better decision-making for inventory management, marketing strategies, and resource allocation within the e-commerce platform.

2.8 E-COMMERCE SALES PREDICTION USING MACHINE LEARNING ALGORITHM

In [8] Karandeep Singh et.al., proposed the implementation of e-commerce system centered on sales prediction employing machine learning techniques. The system likely involves gathering extensive data on various aspects like historical sales, customer behavior, and marketing initiatives. Singh's approach likely includes preprocessing and feature engineering to extract relevant information from this data. He probably employs machine learning algorithms such as regression or neural networks to train predictive models. Once trained, these models can forecast future sales trends, aiding businesses in making informed decisions regarding inventory management, marketing strategies, and resource allocation. Singh's system likely stands to revolutionize e-commerce operations by providing accurate and actionable insights for optimizing sales performance.

2.9 SALES PREDICTION USING MACHINE LEARNING

In [9] Sumanth Mekala et.al., proposed the implementation of Predicting sales in e-commerce using machine learning involves leveraging historical sales data, customer behavior, and various external factors to forecast future sales accurately. Initially, the project entails data collection, cleaning, and preprocessing to ensure data quality. Features such as seasonality, trends, and promotional events are then extracted through feature engineering. Subsequently, a suitable machine learning model is selected and trained on the prepared dataset, considering factors like regression, time series analysis, or neural networks. Model evaluation and optimization are crucial to enhance predictive accuracy. Finally, the developed model is deployed into production, integrated with the e-commerce platform, and regularly monitored for performance. Collaboration between data scientists, domain experts, and stakeholders is vital throughout the project to align objectives and drive actionable insights for informed decision-making in e-commerce sales strategies.

2.10 SALES PREDICTION USING MACHINE LEARNING

In [10] Ashok kumar et.al., proposed the implementation of e-commerce sales prediction using machine learning, the process begins with gathering, cleaning, and preparing data for analysis. This involves incorporating historical sales data, customer behavior patterns, and external variables that could influence sales. Feature engineering is then applied to extract relevant features such as seasonality, trends, and promotional events. The next step involves selecting an appropriate machine learning model, which could range from regression techniques to time series analysis or neural networks, depending on the nature of the data and the prediction task. The model is trained on the prepared dataset and evaluated for accuracy, with optimization techniques applied to enhance performance. Once the model is deemed satisfactory, it is deployed into production, integrated with the e-commerce platform, and monitored regularly for any deviations or changes in performance. Collaboration among data scientists, domain experts, and stakeholders is essential throughout the project to ensure alignment with business objectives and facilitate data-driven decision-making in e-commerce sales strategies.

2.11 SUMMARY

E-commerce sales forecasting with machine learning involves collecting and preprocessing data, including historical sales and external factors. Feature engineering extracts relevant patterns like seasonality and trends. A suitable model, such as regression or neural networks, is selected, trained, and optimized for accuracy. Deployment integrates the model with the e-commerce platform, followed by regular performance monitoring. Collaboration among stakeholders ensures alignment with business goals for informed decision-making.

CHAPTER 3

EXISTING SYSTEM

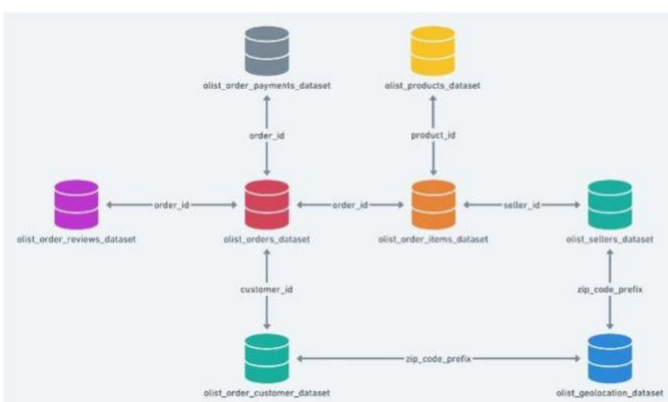
CHAPTER 3

EXISTING SYSTEM

3.1 EXISTING SYSTEM

In the realm of e-commerce, machine learning techniques are extensively utilized for sales prediction. These methods involve the analysis of historical sales data, customer behavior, market trends, and various other factors to forecast future sales accurately. Typically, the existing systems for e-commerce sales prediction leverage algorithms such as linear regression, decision trees, random forests, support vector machines, and neural networks. These algorithms are trained on vast datasets comprising information like past sales records, product attributes, promotional activities, customer demographics, website traffic, and seasonality effects. Moreover, feature engineering plays a crucial role in enhancing the predictive performance of these models. Features like time of day, day of week, special events, and holidays are often incorporated to capture temporal patterns and anomalies. Furthermore, advanced techniques like ensemble learning, time series analysis, and deep learning are being increasingly explored to improve prediction accuracy and robustness. Ensemble methods combine multiple models to produce more accurate predictions, while time series analysis techniques handle data sequences efficiently, capturing trends and seasonal patterns. Deep learning models, particularly recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), excel in capturing complex temporal dependencies and are adept at forecasting sequential data.

3.2 BLOCK DIAGRAM OF THE EXISTING SYSTEM



3.3 ALGORITHM AND METHODOLOGY OF THE EXISTING SYSTEM

3.3.1 ALGORITHM

In e-commerce systems, predicting sales is crucial for inventory management, marketing strategies, and overall business planning. Various machine learning algorithms can be employed for this purpose. One common approach is to use regression techniques, where historical sales data and relevant features are used to train a model to predict future sales. One popular algorithm for sales prediction in e-commerce systems is the Random Forest algorithm. Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and generalize well to new data. In the context of sales prediction, Random Forest can handle complex relationships between features and sales, such as seasonality, promotions, and customer behavior. Another commonly used algorithm is Gradient Boosting, particularly XGBoost or LightGBM. Gradient Boosting builds a series of weak learners (typically decision trees) sequentially, with each subsequent learner focusing on the errors of the previous ones. This iterative process allows Gradient Boosting models to capture complex patterns in the data and make accurate predictions. Additionally, neural networks, especially recurrent neural networks (RNNs) or long short-term memory networks (LSTMs), can be used for sales prediction in e-commerce systems. These models excel at capturing temporal dependencies and patterns in sequential data, making them suitable for tasks such as sales forecasting, where time-series data is involved. Ultimately, the choice of algorithm depends on factors such as the size and complexity of the dataset, the availability of features, and computational resources. Experimentation with different algorithms and model architectures is often necessary to determine the best approach for a specific e-commerce system's sales prediction task.

3.3.2 METHODOLOGY

- Data Collection: Data was obtained from a study conducted at a Brazilian hospital, involving adult patients hospitalized between June 2016 and December 2017. This dataset served as the basis for training and evaluating the ML models.
- Feature Selection: Feature selection techniques were employed to identify the most relevant variables or features from the dataset that could contribute to predicting ADRs. This helps in improving model performance and reducing computational complexity.

- Class Balancing Techniques: Imbalanced datasets, where one class (e.g.,ADR occurrence) is significantly more prevalent than the other, can lead to biased models. Class balancing techniques, such as oversampling or under sampling, were likely applied to address this issue and ensure that the models are trained on balanced datasets.
- Evaluation Metrics: The performance of the ML models was evaluated using five-fold cross-validation. This involves partitioning the dataset into five subsets, training the model on four subsets, and validating it on the fifth subset. This process is repeated five times, with each subset used as the validation set once. Evaluation metrics such as Area under the ROC Curve (AUC), recall, precision, and F-measure were used to assess the models' performance.

3.4 DRAWBACKS OF THE EXISTING SYSTEM

Potential drawback is the need for a large amount of quality data to train the machine learning models effectively. Without enough high-quality data, the accuracy and reliability of the predictions may suffer. Secondly, there's a risk of over fitting the models to the training data, which could result in poor generalization to unseen data. This is especially true if the models are overly complex or if the training data is not representative of the real-world distribution of sales data. Thirdly, the system may require significant computational resources, especially if using complex machine learning algorithms or processing large volumes of data. This could result in high infrastructure costs and may not be feasible for all businesses, particularly smaller ones with limited resources. Another potential drawback is the need for continuous maintenance and updates to keep the models relevant and accurate over time. Markets and consumer behavior can change rapidly, and the models will need to be regularly retrained with new data to stay effective. Additionally, there are ethical considerations surrounding the use of machine learning in e-commerce, such as privacy concerns related to the collection and use of customer data. Ensuring compliance with regulations and maintaining transparency with users about how their data is being used is essential but can be challenging. Finally, the interpretability of machine learning models can be a concern. If the models are too complex, it may be difficult for stakeholders to understand how predictions are being made, which could limit trust in the system and hinder its adoption.

CHAPTER 4

PROPOSED SYSTEM

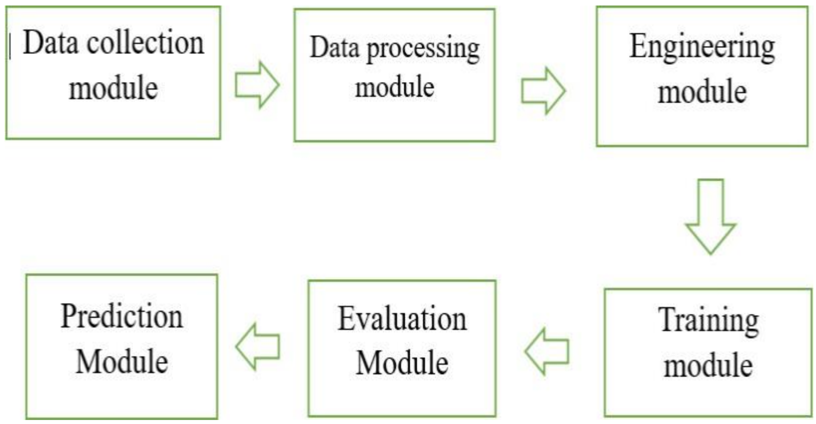
CHAPTER 4

PROPOSED SYSTEM

4.1 PROPOSED SYSTEM

In the increasingly competitive landscape of e-commerce, accurate sales prediction is paramount for businesses to optimize inventory management, marketing strategies, and overall performance.. Our proposed system integrates various machine learning algorithms to analyze historical sales data, customer behavior, product attributes, and external factors such as seasonality, promotions, and economic indicators. The system encompasses several key components Gather historical sales data, including transactional records, customer demographics, product details, and any other relevant information. Cleanse and preprocess the data to handle missing values, outliers, and inconsistencies. Feature engineering to extract meaningful features and transform raw data into a suitable format for modeling. Experiment with a variety of machine learning algorithms such as linear regression, decision trees, random forests, gradient boosting, and neural networks. Deploy the trained models to predict future sales based on input variables such as time period, product category, pricing, and promotional activities. Implement real-time forecasting capabilities to adapt to dynamic market conditions and emerging trends. Evaluate the predictive performance of the models using appropriate metrics such as mean absolute error (MAE), root mean squared error (RMSE), or mean absolute percentage error (MAPE). Continuously monitor model performance and recalibrate as necessary to maintain accuracy over time. Integrate feedback loops to incorporate new data and update models periodically for improved forecasting accuracy. Integrate the sales prediction system into the existing e-commerce platform, enabling seamless access to insights for stakeholders. Provide user-friendly interfaces and visualization tools to facilitate interpretation and decision-making. Ensure scalability and robustness to accommodate increasing volumes of data and user traffic. By implementing this comprehensive system for e-commerce sales prediction, businesses can gain valuable insights into future demand trends, optimize inventory levels, allocate resources efficiently, and enhance overall competitiveness in the digital marketplace.

4.2 BLOCK DIAGRAM OF THE PROPOSED SYSTEM



4.3 ALGORITHM AND METHODOLOGY OF THE PROPOSED SYSTEM

4.3.1 ALGORITHM

One approach to developing a CNN (Convolutional Neural Network) algorithm for sale prediction in an e-commerce system involves leveraging machine learning techniques. CNNs are particularly well-suited for tasks involving image recognition, but they can also be adapted for other types of data, including structured data commonly found in e-commerce systems. To create such an algorithm, start by gathering a dataset consisting of historical sales data along with relevant features such as product attributes, time of sale, customer demographics, and marketing efforts. This dataset would be preprocessed to handle missing values, normalize features, and encode categorical variables. The CNN architecture would then be designed to take advantage of the spatial relationships within the data. For example, if the dataset includes images of products, the CNN could use convolutional layers to extract features from these images. If the data is tabular, the CNN could incorporate 1D convolutional layers to capture patterns within sequences of data. Once the architecture is defined, the model would be trained using the prepared dataset. During training, the model learns to map input features to sales predictions by adjusting its parameters to minimize a chosen loss function, such as mean squared error or mean absolute error. After training, the model would be evaluated on a separate test dataset to assess its performance. Metrics such as mean absolute error, mean squared error, or accuracy could be used to measure how well the model generalizes to unseen data. Finally, the trained model could be deployed within the e-commerce system to make real-time sales predictions. This would involve integrating the model into the system's backend infrastructure, where it can receive input data and return predictions for future sales. Overall, developing a CNN algorithm for sale prediction in an e-commerce system involves data collection, preprocessing, model design, training, evaluation, and deployment, all with a focus on leveraging machine learning techniques to improve prediction accuracy.

4.3.2 METHODOLOGY

Convolutional Neural Networks involves gathering historical data on sales, including factors such as product features, customer demographics, seasonality, and marketing efforts. Data preprocessing is crucial to clean and prepare the data for training the model. Feature engineering involves selecting and transforming relevant features that can help the model learn patterns and make accurate predictions. This may include encoding categorical variables, scaling numerical features, and creating new features based on domain knowledge. the architecture of the CNN model is designed. CNNs are well-suited for analyzing spatial patterns in data, making them effective for tasks like image recognition and, by extension, e-commerce sales prediction. The architecture typically consists of convolutional layers, pooling layers, and fully connected layers. Once the model architecture is defined, it is trained on the preprocessed data. After training, the model's performance is evaluated on a separate validation dataset to ensure that it generalizes well to unseen data. Various metrics such as mean squared error or mean absolute error may be used to assess the model's accuracy. Additionally, the model may be tested on a holdout test dataset to further evaluate its performance. Once the model is trained and validated, it can be deployed into the e-commerce system to make real-time sales predictions. It's essential to monitor the model's performance over time and retrain it periodically with new data to maintain its accuracy and effectiveness.

4.4 ADVANTAGES OF THE PROPOSED SYSTEM

- Improved Data Quality: Through advanced data preprocessing techniques and quality assurance measures, it enhances the reliability and completeness of healthcare datasets, mitigating issues such as missing values and biases.
- Balanced Predictive Performance: By employing sophisticated algorithms that address

class imbalances, it ensures fair and accurate predictions across diverse demographic groups, improving overall model performance and reliability.

- Enhanced Interpretability: Despite leveraging deep learning, it incorporates interpretability methods to provide healthcare professionals with meaningful insights into model decisions, promoting trust and facilitating informed decision-making.
- Robust Generalizability: With rigorous external validation and validation on diverse datasets, it demonstrates robust generalizability across different healthcare settings, reducing the risk of over fitting and ensuring reliable performance in real-world scenarios.
- Ethical Compliance and Fairness: It adheres to ethical and legal guidelines, prioritizing patient privacy, consent, and fairness. By addressing biases in both data and algorithms, it fosters trust and equity in healthcare delivery, ensuring that all patients receive fair and unbiased treatment.

CHAPTER 5

IMPLEMENTATION SETUP

CHAPTER 5

IMPLEMENTATION SETUP

5.1 DATA COLLECTION

In the realm of e-commerce sales forecasting with machine learning, data collection and preprocessing are foundational steps crucial for building accurate predictive models. The Data collection involves gathering relevant information from various sources such as transaction records, customer demographics, website traffic logs, product attributes, and external factors like economic indicators or seasonal trends. These datasets provide the raw material necessary for training a forecasting model.

5.2 DATA PREPROCESSING

Once collected, the data undergoes preprocessing to ensure its quality and suitability for analysis. This includes steps such as cleaning the data to remove errors or inconsistencies, handling missing values through imputation or deletion, and encoding categorical variables into numerical representations. Feature engineering is another important aspect of preprocessing, where new features are created or existing ones are transformed to better capture patterns in the data. For instance, time series data might be aggregated into different time intervals or lag features could be generated to capture temporal dependencies. Normalization or standardization is often applied to scale the features to a similar range, which can improve the performance of certain machine learning algorithms. Additionally, data splitting into training, validation, and testing sets is essential to assess the model's performance on unseen data and prevent over fitting.

5.3 EXPLORATORY DATA ANALYSIS

Gather the relevant data, including sales transactions, product details, customer information, and any other variables that may impact sales. Clean the data to address any inconsistencies, missing values, or outliers. This step ensures the reliability and accuracy of the analysis. Calculate basic statistics such as mean, median, standard deviation, and range for numerical variables like sales volume, price, and customer age. For categorical variables like product category or customer segment, examine frequency distributions. Visualize the data using plots and charts to identify trends, seasonality, and outliers.

5.4 FEATURE SELECTION

Feature selection is a crucial step in machine learning that aims to identify the most relevant features for predicting the target variable. In this analysis, a Random Forest Classifier is employed to determine feature importance. The top selected features and their corresponding importance are extracted and presented. Notably, features such as Reaction Outcome exhibit higher importance values, indicating their significant predictive power in determining adverse reactions. Additionally, the analysis highlights the importance of certain categorical features, including 'Occurrence Country' and 'Seriousness Death', alongside numerical features like 'Patient ID' and 'Age'. The bar plot visualization offers a clear representation of feature importance, facilitating the interpretation of key predictors influencing adverse reaction outcomes. Such insights enable better understanding and refinement of the predictive model, ultimately enhancing its accuracy and efficacy in adverse reaction prediction tasks.

5.5 TRAINING AND TESTING DATASETS

This phase involves feeding the model with historical data, including information such as product attributes, customer demographics, purchase history, seasonal trends, and any other relevant variables. This data serves as the foundation for the model to learn patterns and relationships that can be used to predict future sales. Once the model is trained, it needs to be tested to assess its performance and generalization ability. This is typically done by using a separate dataset, called the test set, which the model hasn't seen during training. The model's predictions are compared against the actual sales data to measure its accuracy, precision, recall, F1 score, and other relevant metrics.

5.6 EVALUATION OF THE MODEL

In this phase, the model's performance is thoroughly evaluated to determine its effectiveness in predicting sales accurately. Various techniques such as cross-validation, confusion matrix analysis, ROC curves, and precision-recall curves may be employed to assess different aspects of the model's performance. Additionally, business-specific metrics like revenue generated, customer retention, and inventory optimization may also be considered in evaluating the model's effectiveness in driving business outcomes.

5.7 FEATURE SELECTION

Feature selection is a crucial step in machine learning that aims to identify the most relevant features for predicting the target variable. In this analysis, a Random Forest Classifier is employed to determine feature importance. The top selected features and their corresponding importance are extracted and presented. Notably, features such as Reaction Outcome exhibit higher importance values, indicating their significant predictive power in determining adverse reactions. Additionally, the analysis highlights the importance of certain categorical features, including 'Occurrence Country' and 'Seriousness Death', alongside numerical features like 'Patient ID' and 'Age'. The bar plot visualization offers a clear representation of feature importance, facilitating the interpretation of key predictors influencing adverse reaction outcomes. Such insights enable better understanding and refinement of the predictive model, ultimately enhancing its accuracy and efficacy in adverse reaction prediction tasks.

5.8 TRAINING AND TESTING DATASETS

This phase involves feeding the model with historical data, including information such as product attributes, customer demographics, purchase history, seasonal trends, and any other relevant variables. This data serves as the foundation for the model to learn patterns and relationships that can be used to predict future sales. Once the model is trained, it needs to be tested to assess its performance and generalization ability. This is typically done by using a separate dataset, called the test set, which the model hasn't seen during training. The model's predictions are compared against the actual sales data to measure its accuracy, precision, recall, F1 score, and other relevant metrics.

5.9 EVALUATION OF THE MODEL

In this phase, the model's performance is thoroughly evaluated to determine its effectiveness in predicting sales accurately. Various techniques such as cross-validation, confusion matrix analysis, ROC curves, and precision-recall curves may be employed to assess different aspects of the model's performance. Additionally, business-specific metrics like revenue generated, customer retention, and inventory optimization may also be considered in evaluating the model's effectiveness in driving business outcomes.

CHAPTER 6

RESULT AND INFERENCES

CHAPTER 6

RESULT AND INFERENCES

6.1 EVALUATION METRICS

- Classification Report: This report provides detailed metrics for each class, including precision, recall, and F1-score. These metrics are useful for understanding the performance of the classifier for each class individually.
- Confusion Matrix**:** The confusion matrix visualizes the performance of the classifier by comparing predicted labels with true labels. It provides insights into the number of true positives, true negatives, false positives, and false negatives.
- Receiver Operating Characteristic (ROC) Curve**:** The ROC curve illustrates the performance of a binary classifier across different thresholds by plotting the true positive rate against the false positive rate. The area under the ROC curve (AUC-ROC) quantifies the overall performance of the classifier. A higher AUC-ROC score indicates better discriminatory power.

6.2 RESULTS

The e-commerce industry is rapidly evolving, driven by the surge in online shopping and advancements in technology. To effectively navigate this dynamic landscape, businesses are increasingly turning to machine learning techniques for sales prediction in their e-commerce systems. By leveraging vast amounts of historical data on customer behavior, product attributes, market trends, and other relevant factors, machine learning models can forecast future sales with remarkable accuracy. These models employ various algorithms such as linear regression, decision trees, random forests, and neural networks to analyze patterns within the data and make predictions. Through continuous learning and adaptation, they can refine their forecasts over time, enabling businesses to optimize inventory management, marketing strategies, pricing decisions, and overall operational efficiency.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

The e-commerce industry is rapidly evolving, driven by the surge in online shopping and advancements in technology. To effectively navigate this dynamic landscape, businesses are increasingly turning to machine learning techniques for sales prediction in their e-commerce systems. By leveraging vast amounts of historical data on customer behavior, product attributes, market trends, and other relevant factors, machine learning models can forecast future sales with remarkable accuracy. These models employ various algorithms such as linear regression, decision trees, random forests, and neural networks to analyze patterns within the data and make predictions.

7.2 FUTURE ENHANCEMENT

In the ever-evolving landscape of e-commerce, enhancing sales prediction through machine learning techniques is paramount for staying ahead of the competition. One promising future enhancement is the integration of advanced recommendation systems. By analyzing user behavior, preferences, and past purchases, these systems can offer personalized product recommendations, thereby increasing the likelihood of sales. Additionally, leveraging natural language processing (NLP) algorithms can enable the system to understand and respond to customer inquiries more effectively, facilitating smoother interactions and potentially boosting sales conversion rates. Furthermore, incorporating sentiment analysis tools can help gauge customer satisfaction and sentiment towards products, allowing for proactive adjustments to marketing strategies or product offerings. Finally, harnessing the power of predictive analytics coupled with big data can enable the system to anticipate trends, identify emerging markets, and optimize inventory management, leading to more accurate sales forecasts and better allocation of resources. By continuously innovating and integrating these advancements, e-commerce platforms can not only enhance sales prediction but also provide a more personalized and seamless shopping experience for customers.

REFERENCES

[1] Mr. Ankush Y. Jondhale, Mr. Mohid T. Khan, Mr. Avinash B. Wagh, Mr. Samarth, Prof. A. S. Kolhe, "*E-Commerce Website Sales Prediction using Machine Learning Algorithm*", Volume 4, Issue 2, International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) March 2024.

[2] Mr. Ankush Y. Jondhale, Mr. Mohid T. Khan, Mr. Avinash B. Wagh, Mr. Samarth, Prof. A. S. Kolhe, "*E-Commerce Website Sales Prediction using Machine Learning Algorithm*", Volume 4, Issue 2, March 2024.

[3] S. G. Chordiya , Mr. Aher Shubham , Ms. Ahire Gayatri , Mr. Kute Sumeet , Ms. Mandale Harshada, "*Sales Prediction System using Machine Learning*", Volume 11, Issue 3, www.ijirmps.org, May-June 2023.

[4] Sumanth Mekala1, S. Ashok Kumar, "*SALES PREDICTION OF E-COMMERCE USING MACHINE LEARNING ALGORITHM EXTREME GRADIENT BOOST IN COMPARISON WITH KNEIGHBOR REGRESSION TO IMPROVE ACCURACY*", www.ijirmps.org 2023.

[5] Garud Akshada Anil, Chavan Ritambara Shankar, Bobade Prachi Santosh, Gorad Akshada Rajendra, Prof. B.D. Thorat, "*SALES FORECASTING USING MACHINE LEARNING TECHNIQUES*",International Research Journal of Modernization in Engineering Technology and Science, 2023.

[6] Prajwal Amrutkar , Shubhangi Mahadik, "*Sales Prediction Using Machine Learning Techniques*", www.ijrpr.com, 2022.

[7] Tingli Feng, Chenming Niu, and Yuchen, "*E-commerce Sales Forecast Method Based on Machine Learning Models*", https://doi.org/10.2991/978-2-494069-31-2_119 2022.

[8] Sunny Khatri, Merin Meleet,"*E-COMMERCE SALES FORECASTING*", Volume: 07 Issue: 05 International Research Journal of Engineering and Technology (IRJET) May 2020.

[9] Abhijeet Patre, Rumit Jain, Pranita Darveshi, Rovina D'Britto, Sanketi Raut, "*E-COMMERCE SALES PREDICTION*", www.jetir.org 2020.

[10] Jiahua Li , Tao Wang, Zhengshi Chen, Guoqiang Luo, "*Machine Learning Algorithm Generated Sales Prediction for Inventory Optimization in Cross-border E-Commerce*" , Vol.1, Issue 1: 62-74, International Journal of Frontiers in Engineering Technology, 2019.

APPENDIX

A.1 SOURCE CODE

```
import pandas as pd

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inling

import seaborn as sns

sns.set()

from catboost import CatBoostRegressor, Pool, cv

from catboost import MetricVisualizer

from sklearn.model_selection import TimeSeriesSplit

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

from scipy.stats import boxcox

from os import listdir

import warnings

warnings.filterwarnings("ignore", category=DeprecationWarning)

warnings.filterwarnings("ignore", category=UserWarning)

warnings.filterwarnings("ignore", category=RuntimeWarning)

warnings.filterwarnings("ignore", category=FutureWarning)

import shap

data = pd.read_csv("/content/data.csv.xlsx", encoding="ISO-8859-1", dtype={'CustomerID': str})
```

```
data.shape
from google.colab import drive
drive.mount('/content/drive')
data.head()
missing_percentage = data.isnull().sum() / data.shape[0] * 100
missing_percentage
data[data.Description.isnull()].head()
data[data.Description.isnull()].CustomerID.isnull().value_counts()
data[data.Description.isnull()].UnitPrice.value_counts()
data[data.CustomerID.isnull()].head()
data.loc[data.CustomerID.isnull(), ["UnitPrice", "Quantity"]].describe()
data.loc[data.Description.isnull()==False, "lowercase_descriptions"] = data.loc[
    data.Description.isnull()==False,"Description"
].apply(lambda l: l.lower())
data.lowercase_descriptions.dropna().apply(
    lambda l: np.where("nan" in l, True, False)
).value_counts()
data.lowercase_descriptions.dropna().apply(
    lambda l: np.where("" == l, True, False)
).value_counts()
data.loc[data.lowercase_descriptions.isnull()==False, "lowercase_descriptions"] = data.loc[
    data.lowercase_descriptions.isnull()==False, "lowercase_descriptions"
].apply(lambda l: np.where("nan" in l, None, l))
```

```
Data=data.loc[(data.CustomerID.isnull()==False)
(data.lowercase_descriptions.isnull()==False)].copy()

data.isnull().sum().sum()

data["InvoiceDate"] = pd.to_datetime(data.InvoiceDate, cache=True)


data.InvoiceDate.max() - data.InvoiceDate.min()

print("Datafile starts with timepoint {}".format(data.InvoiceDate.min()))

print("Datafile ends with timepoint {}".format(data.InvoiceDate.max()))

data.InvoiceNo.nunique()

data["IsCancelled"]=np.where(data.InvoiceNo.apply(lambda l: l[0]=="C"), True, False)

data.IsCancelled.value_counts() / data.shape[0] * 100

data.loc[data.IsCancelled==True].describe()

data = data.loc[data.IsCancelled==False].copy()

data = data.drop("IsCancelled", axis=1)

data.StockCode.nunique()

stockcode_counts = data.StockCode.value_counts().sort_values(ascending=False)

fig, ax = plt.subplots(2,1,figsize=(20,15))

sns.barplot(stockcode_counts.iloc[0:20].index,

stockcode_counts.iloc[0:20].values,

ax = ax[0], palette="Oranges_r")

ax[0].set_ylabel("Counts")

ax[0].set_xlabel("Stockcode")

ax[0].set_title("Which stockcodes are most common?");
```

```
sns.distplot(np.round(stockcode_counts/data.shape[0]*100,2),
kde=False,
bins=20,
ax=ax[1], color="Orange")
ax[1].set_title("How seldom are stockcodes?")
ax[1].set_xlabel("% of data with this stockcode")
ax[1].set_ylabel("Frequency");
def count_numeric_chars(l):
return sum(1 for c in l if c.isdigit())
data["StockCodeLength"] = data.StockCode.apply(lambda l: len(l))
data["nNumericStockCode"] = data.StockCode.apply(lambda l: count_numeric_chars(l))
fig, ax = plt.subplots(1,2,figsize=(20,5))
sns.countplot(data["StockCodeLength"], palette="Oranges_r", ax=ax[0])
sns.countplot(data["nNumericStockCode"], palette="Oranges_r", ax=ax[1])
ax[0].set_xlabel("Length of stockcode")
ax[1].set_xlabel("Number of numeric chars in the stockcode");
data.loc[data.nNumericStockCode < 5].lowercase_descriptions.value_counts()
data = data.loc[(data.nNumericStockCode == 5) & (data.StockCodeLength==5)].copy()
data.StockCode.nunique()
data = data.drop(["nNumericStockCode", "StockCodeLength"], axis=1)
data.Description.nunique()
description_counts = data.Description.value_counts().sort_values(ascending=False).iloc[0:30]
plt.figure(figsize=(20,5))
```

```
sns.barplot(description_counts.index, description_counts.values, palette="Purples_r")

plt.ylabel("Counts")

plt.title("Which product descriptions are most common?");

plt.xticks(rotation=90);

def count_lower_chars(l):

return sum(1 for c in l if c.islower())

data["DescriptionLength"] = data.Description.apply(lambda l: len(l))

data["LowCharsInDescription"] = data.Description.apply(lambda l: count_lower_chars(l))

fig, ax = plt.subplots(1,2,figsize=(20,5))

sns.countplot(data.DescriptionLength, ax=ax[0], color="Purple")

sns.countplot(data.LowCharsInDescription, ax=ax[1], color="Purple")

ax[1].set_yscale("log")

data.loc[data.nNumericStockCode < 5].lowercase_descriptions.value_counts()

data = data.loc[(data.nNumericStockCode == 5) & (data.StockCodeLength==5)].copy()

data.StockCode.nunique()

data = data.drop(["nNumericStockCode", "StockCodeLength"], axis=1)

data.Description.nunique()

description_counts = data.Description.value_counts().sort_values(ascending=False).iloc[0:30]

plt.figure(figsize=(20,5))

sns.barplot(description_counts.index, description_counts.values, palette="Purples_r")

plt.ylabel("Counts")

plt.title("Which product descriptions are most common?");

plt.xticks(rotation=90);
```

```
lowchar_counts = data.loc[data.LowCharsInDescription > 0].Description.value_counts()
plt.figure(figsize=(15,3))
sns.barplot(lowchar_counts.index, lowchar_counts.values, palette="Purples_r")
plt.xticks(rotation=90);
def count_upper_chars(l):
return sum(1 for c in l if c.isupper())
data["UpCharsInDescription"] = data.Description.apply(lambda l: count_upper_chars(l))
data.loc[data.UpCharsInDescription <=5].Description.value_counts()
data = data.loc[data.UpCharsInDescription > 5].copy()
dlength_counts = data.loc[data.DescriptionLength < 14].Description.value_counts()

plt.figure(figsize=(20,5))
sns.barplot(dlength_counts.index, dlength_counts.values, palette="Purples_r")
plt.xticks(rotation=90);
data.StockCode.nunique()
data.groupby("StockCode").Description.nunique().sort_values(ascending=False).iloc[0:10]
data.loc[data.StockCode == "23244"].Description.value_counts()
data.CustomerID.nunique()
#plt.xticks(rotation=90);
data.Country.nunique()
country_counts = data.Country.value_counts().sort_values(ascending=False).iloc[0:20]
plt.ylabel("Counts")
plt.title("Which countries made the most transactions?");
```

```
plt.xticks(rotation=90);

plt.yscale("log")

data.loc[data.Country=="United Kingdom"].shape[0] / data.shape[0] * 100

data["UK"] = np.where(data.Country == "United Kingdom", 1, 0)

data.loc[data.UnitPrice == 0].sort_values(by="Quantity", ascending=False).head()

data.loc[data.UnitPrice == 0].sort_values(by="Quantity", ascending=False).head()

#model.show_importances(kind=None)
```

A.2 SCREENSHOTS

A.2.1 Data collection

1	InvoiceNo	StockCode	Descriptio	Quantity	InvoiceDat	UnitPrice	CustomerI	Country
2	536365	85123A	WHITE HA	6	########	2.55	17850	United Kingdom
3	536365	71053	WHITE ME	6	########	3.39	17850	United Kingdom
4	536365	84406B	CREAM CU	8	########	2.75	17850	United Kingdom
5	536365	84029G	KNITTED U	6	########	3.39	17850	United Kingdom
6	536365	84029E	RED WOO	6	########	3.39	17850	United Kingdom
7	536365	22752	SET 7 BAB	2	########	7.65	17850	United Kingdom
8	536365	21730	GLASS STA	6	########	4.25	17850	United Kingdom
9	536366	22633	HAND WA	6	########	1.85	17850	United Kingdom
10	536366	22632	HAND WA	6	########	1.85	17850	United Kingdom
11	536367	84879	ASSORTED	32	########	1.69	13047	United Kingdom
12	536367	22745	POPPY'S PI	6	########	2.1	13047	United Kingdom
13	536367	22748	POPPY'S PI	6	########	2.1	13047	United Kingdom
14	536367	22749	FELTCRAFT	8	########	3.75	13047	United Kingdom
15	536367	22310	IVORY KNI	6	########	1.65	13047	United Kingdom
16	536367	84969	BOX OF 6 A	6	########	4.25	13047	United Kingdom
17	536367	22623	BOX OF VI	3	########	4.95	13047	United Kingdom
18	536367	22622	BOX OF VI	2	########	9.95	13047	United Kingdom
19	536367	21754	HOME BUI	3	########	5.95	13047	United Kingdom
20	536367	21755	LOVE BUIL	3	########	5.95	13047	United Kingdom
21	536367	21777	RECIPE BO	4	########	7.95	13047	United Kingdom
22	536367	48187	DOORMAT	4	########	7.95	13047	United Kingdom
23	536368	22960	JAM MAKI	6	########	4.25	13047	United Kingdom
24	536368	22913	RED COAT	3	########	4.95	13047	United Kingdom
25	536368	22912	YELLOW C	3	########	4.95	13047	United Kingdom
26	536368	22914	BLUE COA	3	########	4.95	13047	United Kingdom
27	536369	21756	BATH BUIL	3	########	5.95	13047	United Kingdom
28	536370	22728	ALARM CL	24	########	3.75	12583	France
29	536370	22727	ALARM CL	24	########	3.75	12583	France

Fig.A.1 Data collection

The above mentioned figure depicts within this kernel we will analyse sales data of an UK online retailer. As storage area may be expensive and fast delivery on time is important to prevail over the competition we like to help the retailer by predicting daily amounts of sold products.

A.2.2 EDA analysis

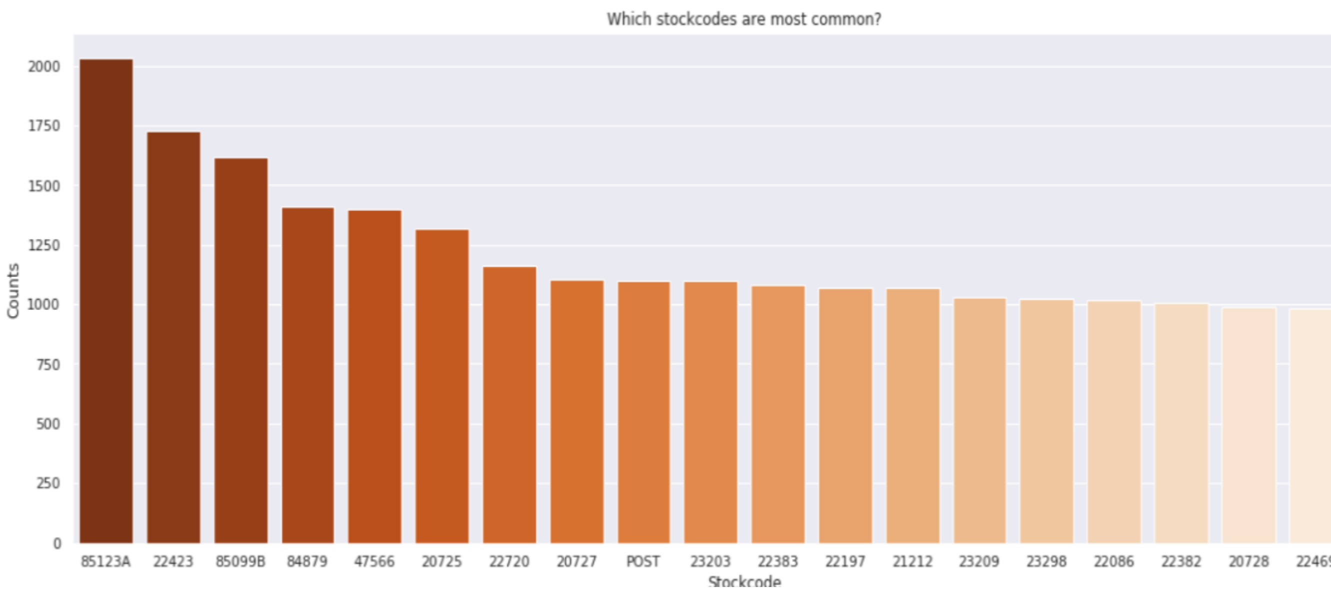


Fig.A.2 EDA analysis on unique stockcode

The above mentioned figures depicts the total count of the user.

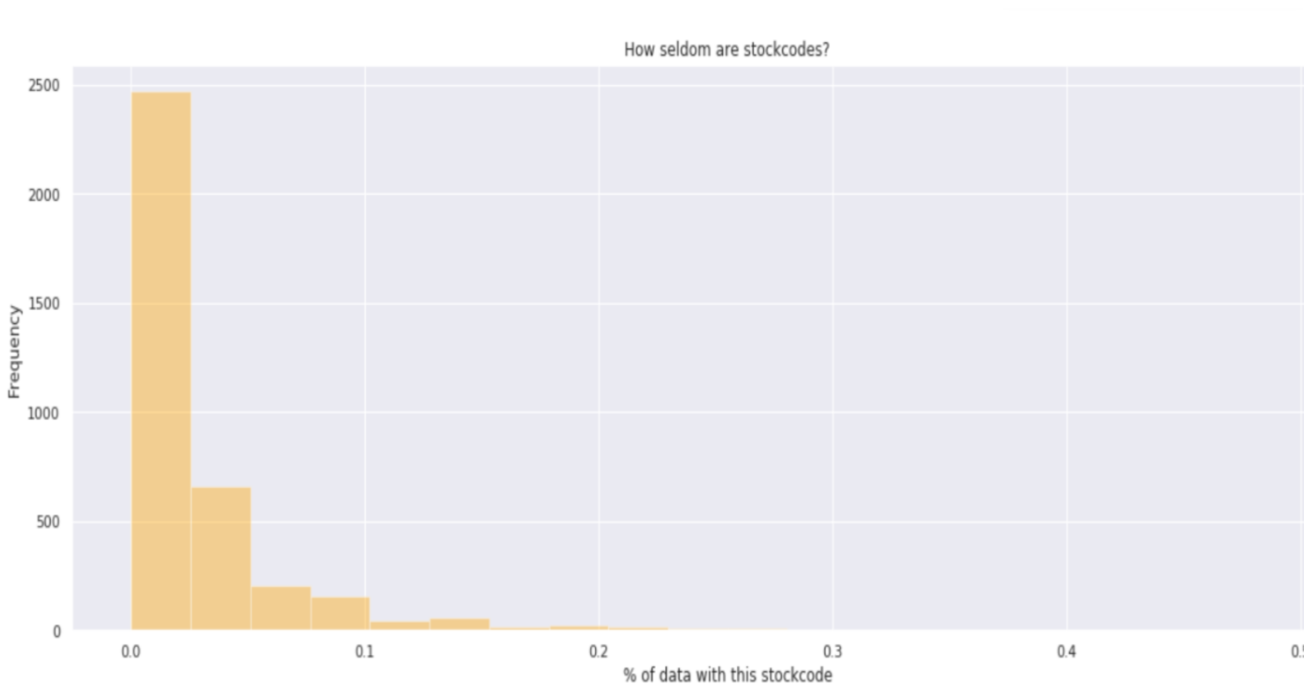


Fig.A.3 EDA analysis seldom are stockcodes

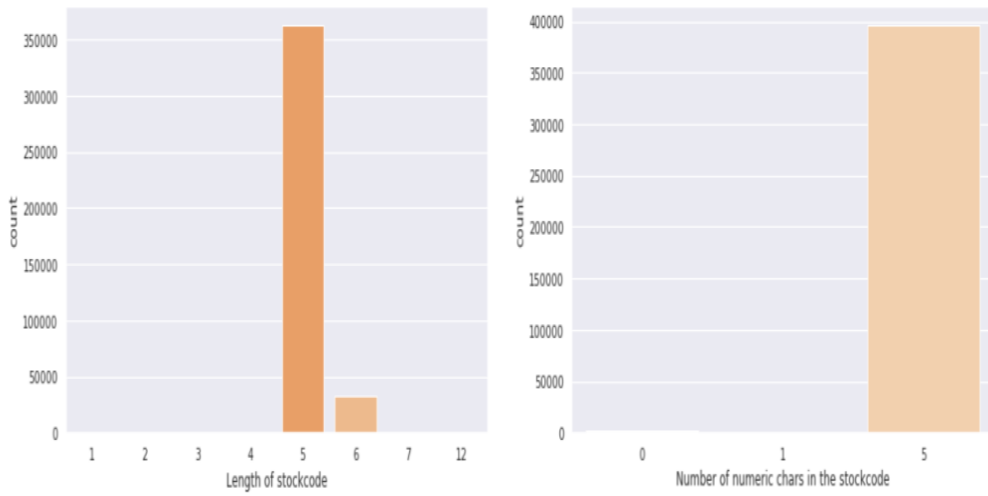


Fig.A.4 EDA analysis on number of numeric chars in the stockcode

The above mentioned figures depicts the Number of stock available for chairs in the stockcode.

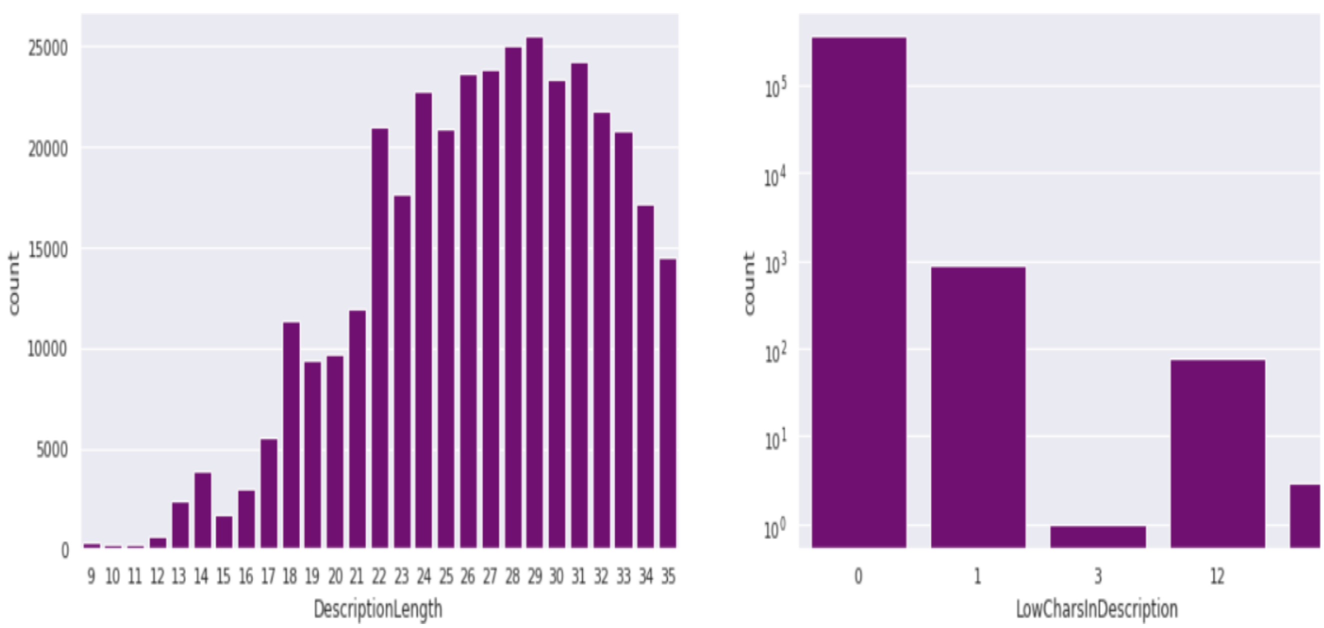


Fig.A.5 EDA analysis on similar product type

The above mentioned figures depicts the EDA analysis on the retailer sells various different kinds of products

A.2.3 Feature selection

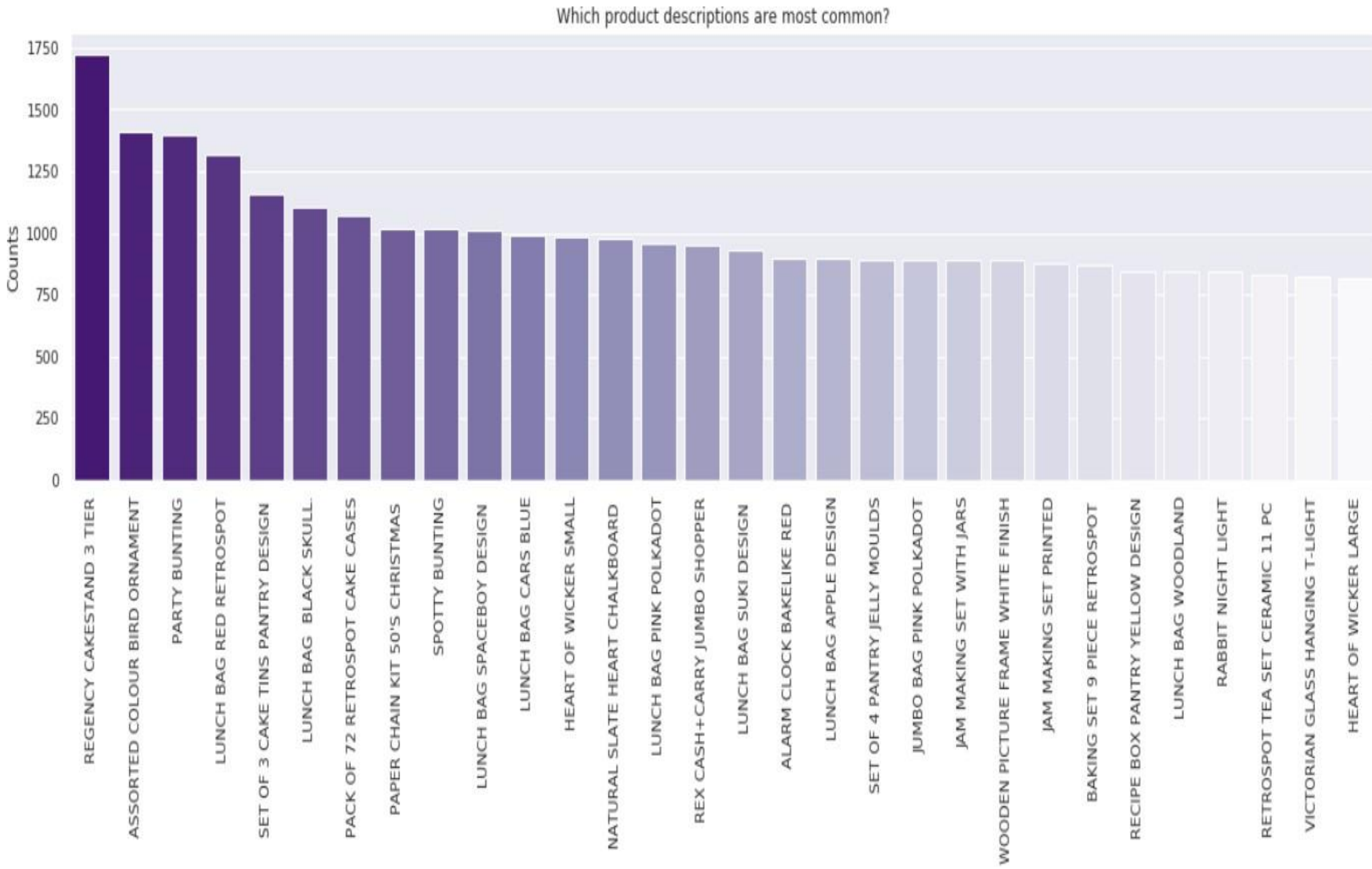


Fig.A.6 Feature selection over the dataset

The above mentioned figures depicts the feature selection on the dataset which will involves identifying and choosing the product which achieve the higher sale point.

A.2.4 Training and testing the model

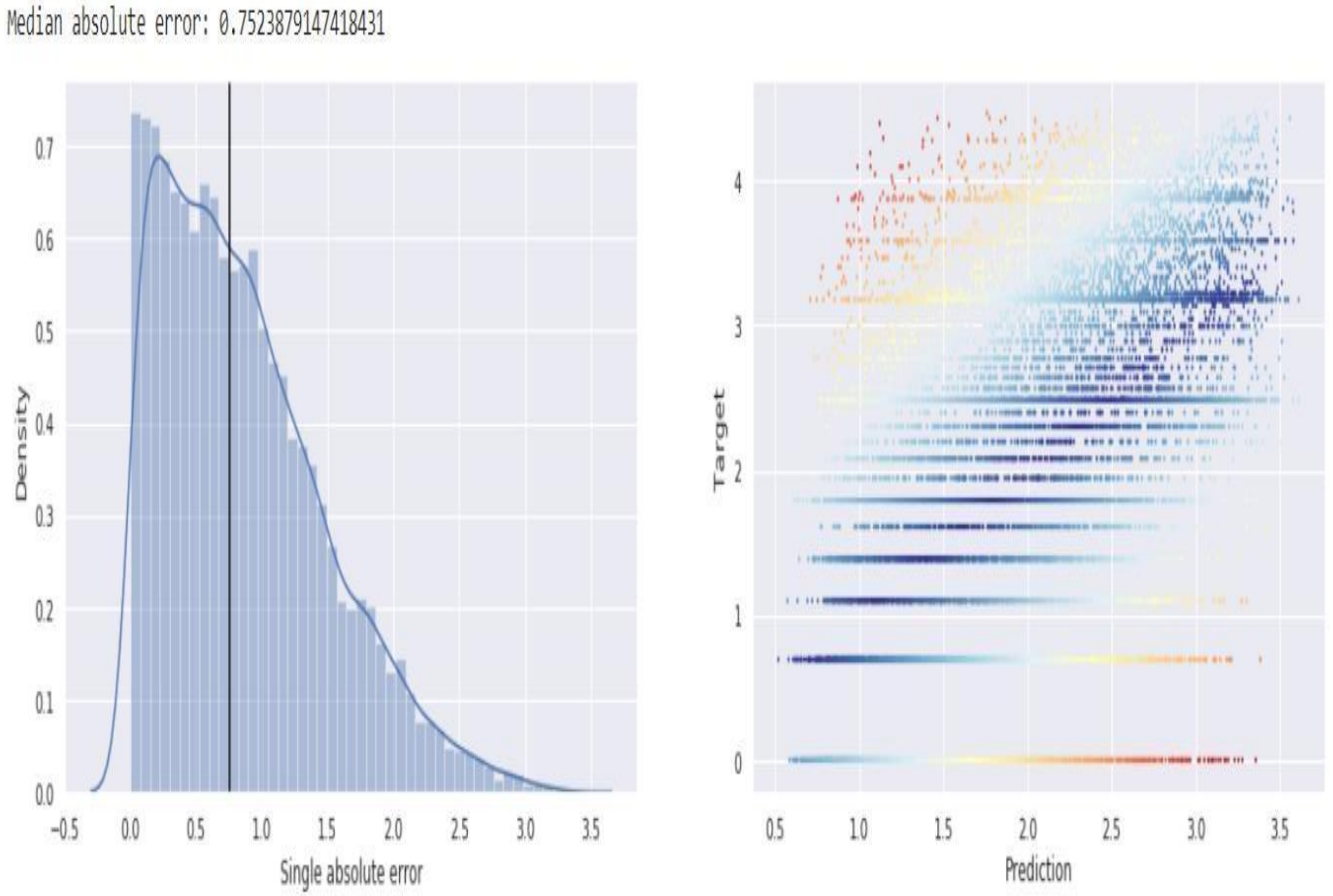


Fig.A.7 accuracy of the dataset

The above mentioned figures depicts the epoch vs accuracy plot, the accuracy of the model is visualized over successive training epochs, indicating how well the model is learning from the training data over time.

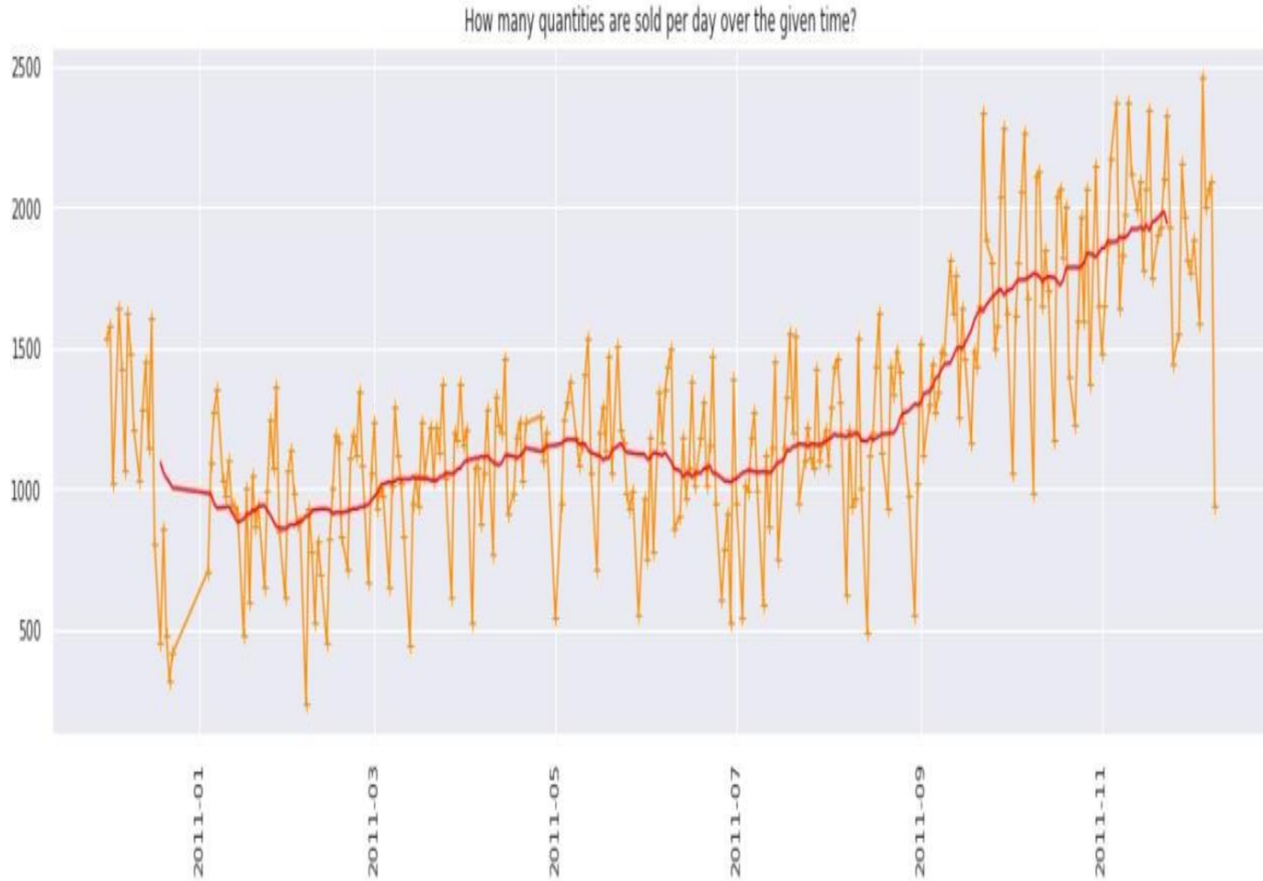


Fig.A.8 loss of the dataset

The above mentioned figures depicts the loss vs epoch plot, the loss function value, representing the discrepancy between predicted and actual values, is depicted across epochs, providing insights into the convergence and stability of the training process.

A.2.5 Evaluating the model

	StockCode	Year	Month	Week	Weekday	Date	Quarter	Dayofyear	Day	Quantity	Revenue	ProductType	KnownStockCodeUnitPriceMedian	KnownStocl
0	10002	2010	12	48	2	2010-12-01	4	335	1	4.094345	51.00	9	0.85	
1	10125	2010	12	48	2	2010-12-01	4	335	1	0.693147	1.70	9	0.85	
2	10133	2010	12	48	2	2010-12-01	4	335	1	1.609438	4.25	14	0.42	
3	16014	2010	12	48	2	2010-12-01	4	335	1	2.302585	4.20	23	0.42	
4	16016	2010	12	48	2	2010-12-01	4	335	1	2.302585	8.50	9	0.85	

Fig.A.9 Classification report of the trained dataset

The above mentioned figures depicts the classification report provides a comprehensive summary of the model's performance, including precision, recall, and F1-score for each class, along with accuracy and support values.

Fig.A.10 Confusion matrix of the trained dataset

CERTIFICATION

udemy

Certificate no: UC-83fa9385-a37a-46b4-8f0e-dabb46da20af
Certificate url: ude.my/UC-83fa9385-a37a-46b4-8f0e-dabb46da20af
Reference Number: 0004

CERTIFICATE OF COMPLETION

Fundamentals of IoT (Internet of Things)

Instructors **Harish Kumar Maheshwari**

HARSHA PRADHA G

Date **May 13, 2024**
Length **34 total mins**

udemy

Certificate no: UC-3e3fffb5-7155-4490-a8
Certificate url: ude.my/UC-3e3fffb5-7155-4490-a8
Referen

CERTIFICATE OF COMPLETION

Fundamentals of IoT (Internet of Things)

Instructors **Harish Kumar Maheshwari**

Haripriya C

Date **May 15, 2024**
Length **34 total mins**

Certificate no: UC-83fa9385-a37a-46b4-8f
Certificate url: ude.my/UC-83fa9385-a37a-46b4-8f
Referen

CERTIFICATE OF COMPLETION

Fundamentals of IoT (Internet of Things)

Instructors **Harish Kumar Maheshwari**

Haripriya S

Date **May 13, 2024**
Length **34 total mins**