JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EVALUATION OF MACHINE LEARNING BASED PREDICTION SYSTEMS FOR COMPARING ONLINE PRODUCT PRICES

S. Amirtharaj[1] and K. Shiva Varshini[2]
[1]Associate Professor, [2]PG Student,
Department of Computer Applications, Mepco Schlenk Engineering College, Sivakasi, India

Abstract *:* In today's dynamic e-commerce world, consumers are faced with difference in product prices across multiple platforms, so there is a need for creating automated price comparison tools for making the user purchase more reliable. This paper presents a comprehensive system that extracts and compares product prices from various e-commerce platforms, specifically Amazon and Flipkart. To further enhance decision-making capabilities, the system includes advanced machine learning (ML) techniques for forecasting future price trends. By enabling predictive insights, users are empowered to make informed decisions regarding optimal purchasing times. The study investigates and compares the performance of several regression-based ML models including Linear Regression, Decision Tree Regressor, Random Forest Regressor, and XGBoost based on their accuracy and predictive reliability. The experimental evaluations of these models demonstrate that the integration of predictive analytics into price comparison systems significantly enhance user decision-making, leading to smarter, data-driven shopping experiences.

IndexTerms **- Web Scraping, Machine Learning, Price Forecasting, E-commerce Intelligence, XGBoost, Random Forest, Price Comparison.**

I. INTRODUCTION

In recent times, the rise of e-commerce are huge which changed the way people purchase things, with various e-commerce platforms are becoming the most preferred choice for purchasing wide range of products. Online shopping provides user with the convenience with availability of multiple sellers which allows the user to explore many options before making a purchase. But in online shopping, there is one significant challenge is that variation in price of the same product across different platforms. The same product may have different prices on various websites which makes the users to spend more amount of time to find the best deal for the product to be purchased across the various websites.

This manual process will be not only time consuming for the user but also inefficient when dealing with frequent change in price by discounts, various pricing strategies and limited offer during the festival season. As the result there will be need for smart way to tackle this by automating the comparison process and help users make more smart purchasing decisions.

To overcome to this challenge, the proposed system use web scraping techniques to extract real time data product price from the e-commerce platforms and it integrates machine learning algorithms to analyze the price of the product over the time and predict the future trends. This will combine the current price of the product with the predicted trends, the system offers timely recommendation on whether to buy a product immediately or wait for price drop in future.

This paper is to compares the performance of several ML models—including Linear Regression, Decision Tree Regressor, Random Forest Regressor, and XGBoost to find which algorithm will be effective for the price forecasting. This is to determine which algorithm performs best in predicting future prices thus enhancing the system's ability to provide accurate, purchase recommendations to users. The final outcome of this study identifies the most effective algorithm based on experimental results and performance metrics which be more effective to be used in the system to make the user experience more easy.

II. LITERATURE SURVEY

An application that compares the best price of goods from various sources which helps the user to make the purchasing decisions to make the user save money is proposed in [1]. In this paper, the user will provide the name of the goods to be purchased and it will list the details of the product in various platforms which done by webscraping and selenium web automation, and python modules were used for data extraction.

A clear view of the role of web scraping is presented in [2] making online shopping easier. It elaborates how price comparison websites use web scraping to gather product data, compare prices, and offer insights to consumers. The study explains the use of tools like BeautifulSoup, Flask, Selenium, Node.js, Next.js, and MongoDB for easy data extraction, processing, and visualization. By discovering hidden pricing strategies and market trends on e-commerce platforms, the work makes the consumers to make smarter, more informed purchase decisions.

A web application developed using python, flask, HTML, CSS and Javascript developed in [3] to make informed purchase decisions. This utilize the web scraping predictive capabilities the application uses historical data in determining both fit attributes

of products and how much they are priced hence it performs comparisons in various platforms seamlessly among various ecommerce platforms.

The review in [4] aims to provide an updated literature survey about the most advanced Web Scraping techniques for better understanding for scholars with helpful knowledge on how to mine online data effectively. This paper starts with presenting the basic design of a web scraper and the applications of web scraping in various sectors and areas. Next, the different Web scraping methods and Web scraping technologies are presented. Finally, a procedure to develop Web scraping with various tools is proposed.

The study in [5] gives the idea about machine learning algorithms linear regression, Random Forest, Decision Tree and XGBoost how these algorithms were used for the forecasting of a particular thing.

The research work in [6] analyzes and compares Random Forest and XGBoost algorithms, it gives the overview of evolution of both the algorithms and study of facets like time complexity, precision, and reliability. It gives full insight of performance metrics, such as the F1-score, Recall, Precision, Mean Squared Error.

The study in [7] focuses on the Linear Regression, Ridge Regression, Support Vector Regression, Lasso Regression. These algorithms are used to implement these models and evaluate the predictions based on certain metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MEA), R2, and Adjusted R2.

The creation of an Automated Product Price Comparison (APPC) system [8] using Python, focusing on web scraping, data cleaning, and analysis. It also integrates machine learning to predict price trends. It addresses key challenges like inconsistent data formats and constantly changing prices, and offers solutions to make the system more effective and reliable.

The study compares SVR, RNN, and LSTM models for predicting market prices using the NGX All-Share Index. LSTM outperformed the others, especially with a 60-day input window, achieving a high R^2 score due to its ability to capture long-term patterns and handle temporal dependencies. SVR struggled with sudden market changes, while RNN faced the vanishing gradient problem in [9]. The use of technical indicators and the optimization framework further improved LSTM's performance, making it the most accurate and reliable model in this analysis.

The performance com[parison [10] of different classification, regression, and clustering algorithms. These algorithms identify patterns and relationships within the data, resulting in precise predictions. There is a growing importance of predictive algorithms in various domains such as finance, healthcare, marketing, weather forecasting, E-commerce, etc. The system enables data scientists and practitioners to make informed decisions when selecting appropriate models for their specific applications.

A system using natural language processing as the AI system to analyze the product evaluations and description to understand the user needs is presented in [11]. The aim of the paper is to develop dynamic and flexible that predicts the price and provides the price comparisons for clothes. It helps the online shoppers a easy way at lower price.

III. PROPOSED WORK

The proposed system combines real-time web scraping with Machine Learning. Users provide a product name, and the system fetches prices from various e-commerce websites. It then uses trained ML models to predict price trends and suggest optimal purchasing period when will the product price will be lower which will benefit the user to save up their money. The data is collected through web scraping techniques, using libraries such as BeautifulSoup and Scrapy, which extract product prices with the date and product name. The system captures price variations on a daily, weekly, and monthly basis, helping to identify underlying patterns, such as festival trends, discounts, and price hikes. This allows the system to understand the impact of difference on prices of the product during different time which will help to forecast future price based on historical sales patterns.

The forecasting mechanism relies on machine learning algorithms which are Linear Regression, Decision Tree, Random Forest, XGBoost. These algorithms are trained on the collected data and evaluated using key performance metrics such as Mean Absolute Error, Root Mean Squared Error and R^2 Score. The system then uses these models to forecast future price trends, which help users to identify the best time to purchase a product at its lowest price, especially during sales events or festival season were the price will be low compare to normal days. By integrating historical price data, sales event information, discount during festival seasons the system provides personalized price predictions and recommendations which be ideal time to buy the product. This allows users to save both time and money by choosing the best time to purchase products based on predicted trends.

IV. SYSTEM ARCHITECTURE

The proposed system is designed to help users easily find the best time and place to buy products in online. The system architecture is illustrated in figure 1. The process begins when a user types in the name of a product they want to search for in the system. This input is sent to a web scraper, which automatically browses e-commerce websites such as Amazon and Flipkart to collect the latest price and product details. The information gathered is then saved in a structured format, typically as a CSV file, which makes it easy to organize and use later. This data is also stored in a database or storage system so that historical pricing can be tracked over time.

At the same time, the collected data is analyzed using a comparison algorithm using machine learning. This algorithm either directly compares the current prices across different websites or uses machine learning models to forecast future price drops based on trends. The purpose of this is to help users decide whether they should buy the product now or wait for a better deal. In the end, the system displays the most relevant and helpful information to the user either showing the lowest current price or recommending the best time to buy making the whole shopping experience smarter, faster, and more cost-effective.

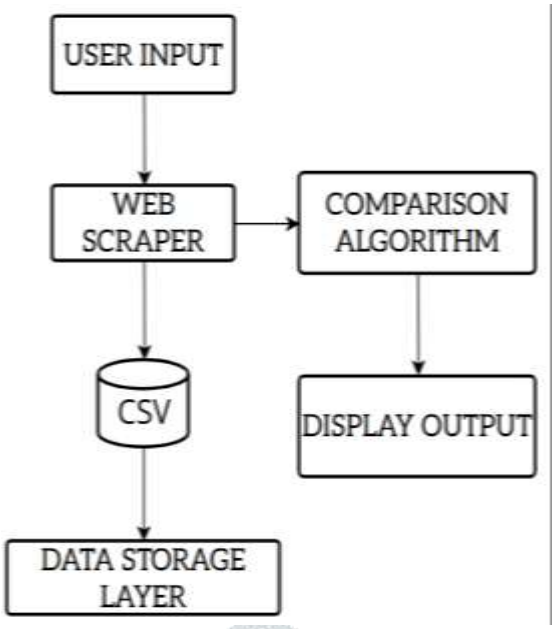

Fig. 1 System Architecture

V. METHODOLOGY

The development of this price comparison and prediction system follows a structured methodology aimed at delivering accurate and timely purchasing suggestions. The process begins with web scraping, where tools like BeautifulSoup or Scrapy are used to extract historical and current product price data from major e-commerce platforms such as Amazon and Flipkart. This data includes not only the product name and price but also date stamps, allowing the system to track how prices fluctuate over time. Once collected, the data is cleaned and organized into a structured format, typically a CSV file or stored directly in a database, enabling easy access and manipulation for training and analysis.

Next, this structured dataset is used to train various machine learning models, including algorithms such as Linear Regression, Decision Tree Regressor, Random Forest Regressor, and XGBoost. These models learn from historical pricing patterns to understand how prices change over time, identifying trends such as discounts during festive sales or gradual price drops. Once trained, the models are capable of taking the latest pricing data and generating price forecasts for selected products.

Based on these predictions, the system provides intelligent recommendations to the user. If the model predicts that a product's price is likely to drop soon, the system suggests the user "Wait for Drop." On the other hand, if the prediction indicates a potential price hike or that the current price is the lowest expected, the user is advised to "Buy Now." This end-to-end methodology ensures that users receive not just static price comparisons, but smart, data-driven guidance to help them make better, more cost-effective shopping decisions.

VI. DATA DESCRIPTION

The dataset utilized for this study includes product pricing information gathered from two major e-commerce platforms, Amazon and Flipkart. The dataset covers a variety of products, including electronics, home appliances, and accessories, with pricing recorded on specific dates during different events such as Diwali Sale, New Year Offer, and regular Normal days. Each record in the dataset contains the following columns: date, product platform, price and event. The data spans across several months and includes a diverse set of products such as smartphones, laptops, air conditioners, washing machines, and more. Notably, the prices for the same products are often different across the two platforms during sale events, reflecting the discounts or seasonal demand changes. The dataset was preprocessed to handle missing values and outliers, and features such as the date were used to extract additional temporal information to analyze trends over time. This dataset offers insights into how prices fluctuate based on events and the platform, providing a rich foundation for applying machine learning algorithms for price prediction.

VII. COMPARATIVE STUDY OF PREDICTION ALGORITHMS

To evaluate the effectiveness of different machine learning models in predicting product prices during various events, four popular algorithms - Linear Regression, Decision Tree, Random Forest and XGBoost were tested and compared based on MAE, RMSE and R^2 Score**.**

7.1 Comparison of Prediction Algorithms

Table 7.1: Comparison Statics

Algorithm	MAE	RMSE	R^2 Score	Remarks
Linear Regression	100.08	118.55	0.316	Weak performance on complex data
Decision Tree	79.97	99.55	0.518	Decent, tends to overfit

Random Forest	68.40	83.27	0.663	Good balance of fit
XGBoost	62.23	78.49	0.700	Best overall performer

A comparison of the performance of the ML based Prediction algorithms is presented in table 7.1. The results indicate that XGBoost performed the best, achieving the lowest MAE of **62.23**, RMSE of **78.49**, and the highest R^2 Score of **0.70**, suggesting a stronger ability to capture price variation. In contrast, Linear Regression showed the weakest performance with an MAE of **100.08** and R^2 Score of **0.31**. The Decision Tree and Random Forest models performed moderately well, with the Random Forest achieving a good balance between accuracy and generalization.

VIII. RESULTS AND DISCUSSION

The models were assessed using three standard regression metrics: Mean Absolute Error, Root Mean Squared Error, and R^2 Score. The results, as illustrated in the individual evaluation charts, highlight distinct differences in performance, particularly under real-world pricing scenarios influenced by seasonal events and platform-based variability.

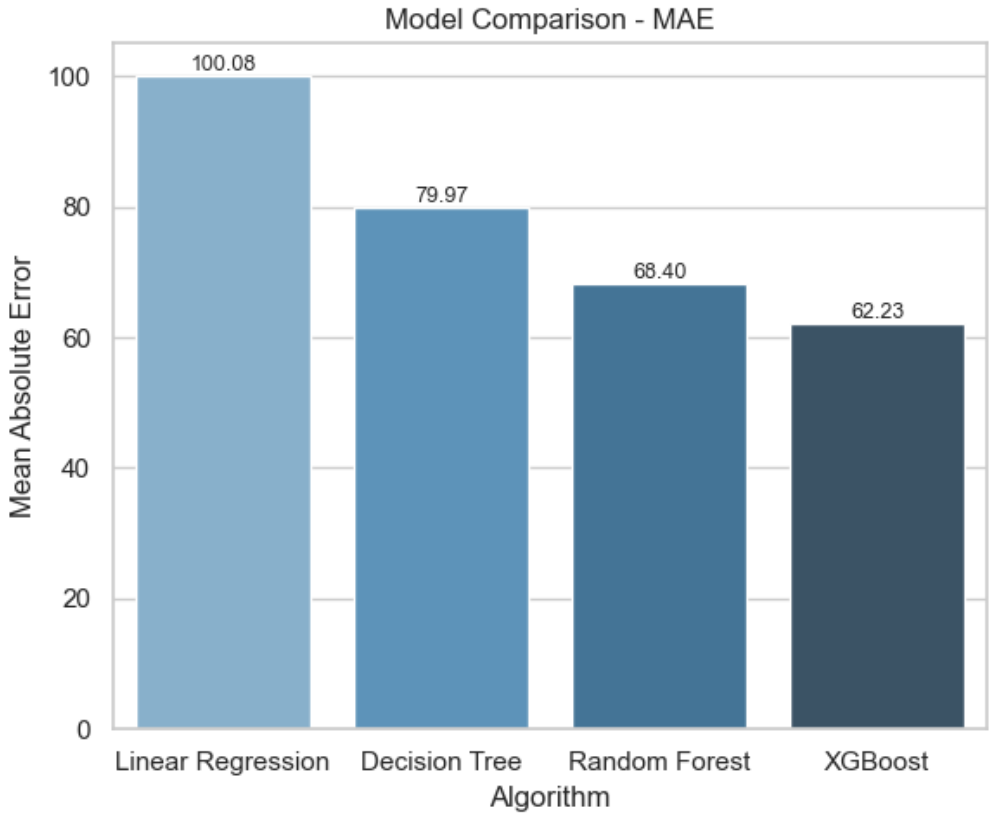

Fig. 2 MAE Comparison of Prediction Models

As shown in the MAE comparison graph presented in Fig 2, XGBoost achieved the lowest error (62.23), followed by Random Forest (68.40). Linear Regression had the highest MAE (100.08), indicating poor performance in capturing price fluctuations.

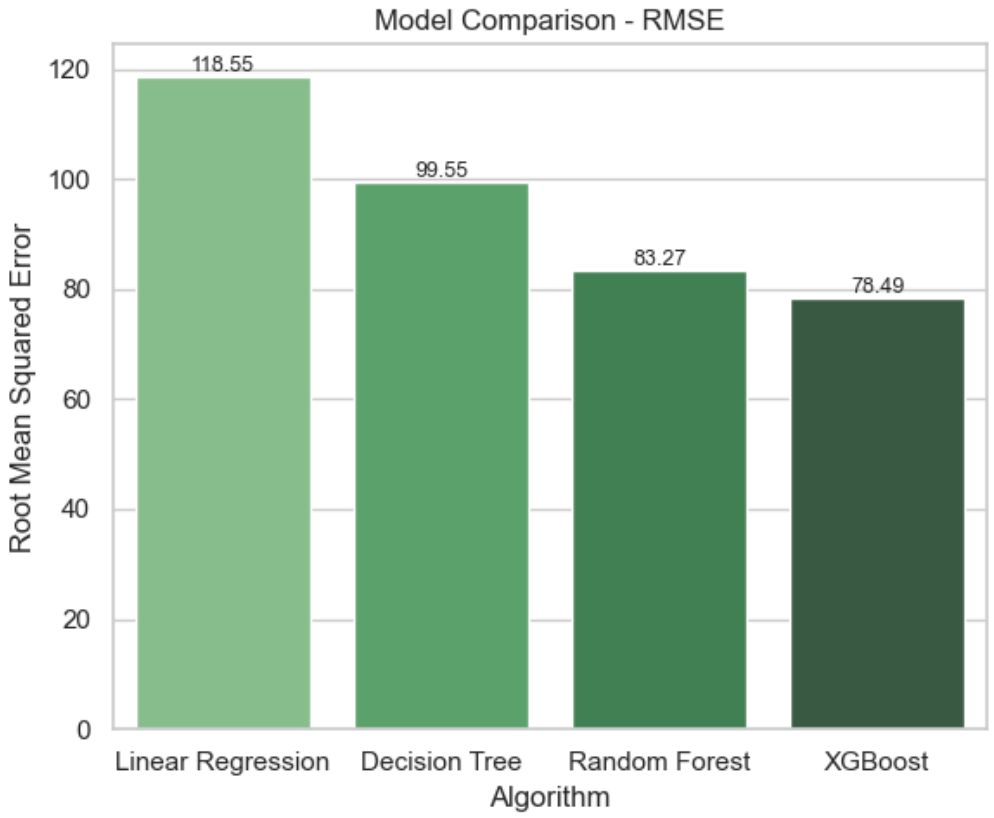

Fig. 3 RMSE Comparison of Prediction Models

Fig 3 presents the comparison of RMSE for the various prediction models presented in this paper.In this scenario also, XGBoost performs best with the lowest RMSE (78.49), suggesting fewer large prediction errors. Random Forest followed with 83.27, while Linear Regression showed the largest RMSE (118.55), making it the least robust.

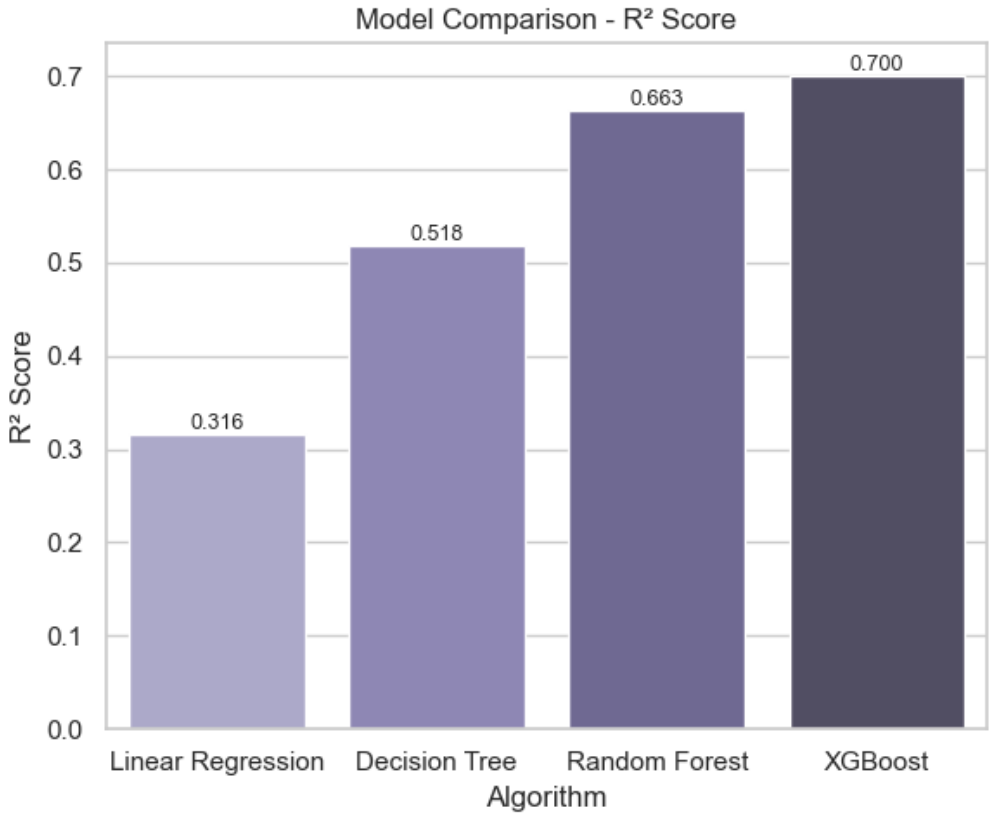

Fig. 4 R^2 Score Comparison of Prediction Models

Fig 4 presents the comparison of R2 scorefor the prediction models whose performances are studied in this paper. XGBoost achieved the highest R^2 score (0.700), explaining 70% of the price variance. Random Forest closely followed with 66.3%, while Linear Regression lagged at 31.6%, indicating a bad fit for the dataset.

IX. CONCLUSION

This study set out to determine the most effective machine learning model for use in a predictive price comparison system. Based on performance evaluation, XGBoost emerged as the best-performing algorithm across MAE, RMSE, and R^2 metrics, validating its strength in modelling complex pricing trends. Random Forest also showed competitive performance. The results highlight that model selection must align with data characteristics; simpler models like Linear Regression are not well-suited for complex datasets as there were more fluctuations. Future enhancements could involve integrating more features such as product ratings, brand popularity, or user demand indicators.

X. ACKNOWLEDGEMENT

The authors are thankful to the host institution for providing the necessary infrastructure and facilities to carry out this research work.

REFERENCES

[1] Varun, K., Rajesh, P. Dileep, P., Ganesh, M., Suneel, P. and Satish, B.S.V. 2003. Price Comparison for Online Shopping. International Journal of Innovative Science and Research Technology, 8(12).

[2] Martina D'Souza, Soham Desai, Dhruv Agrawal and Falguni Joshi, 2024. Web Scraping based Product Comparison Model for E-Commerce Website. Journal of Emerging Technologies and Innovative Research.

[3] Bi Bi Hajira Khanum and Raghavendra, S.P. 2024. Price Comparison using Web-scrapping and Data Analysis, JNNCE Journal of Engineering & Management, Special Edition-2.

[4] Chaimaa Lotfi, Swetha Srinivasan, Myriam Ertz and Imen Latrous. 2021.Web Scraping Techniques and Applications: A Literature Review, SCRS Conference Proceedings on Intelligent Systems.

[5] Muhammet Mustafa Gökçe and Erkan Duman. 2022.Performance Comparison of Simple Regression, Random Forest and XGBoost Algorithms for Forecasting Electricity Demand, 3rd International Informatics and Software Engineering Conference (IISEC)(IEEE).

[6] Sana Fatima and Ayan Hussain and Sohaib Bin Amir and Syed Haseeb Ahmed and Syed Muhammad Huzaifa Aslam. 2023.XGBoost and Random Forest Algorithms: An In Depth Analysis,Pakistan Journal of Scientific Research.

[7] Abhishek Tatachar, V. 2021.Comparative Assessment of Regression Models Based On Model Evaluation Metrics, International Research Journal of Engineering and Technology.

[8] Praveen Kumar. 2025. Automated Product Price Comparison using Python, Nanotechnology Perceptions,21(1).

[9] Olamilekan Shobayo, Sidikat Adeyemi-Longe, Olusogo Popoola and Obinna Okoyeigbo. 2025. A Comparative Analysis of Machine Learning and Deep Learning Techniques for Accurate Market Price Forecasting.

[10] Gupta, S., Kishan, B. and Gulia, P. 2024. Comparative Analysis of Predictive Algorithms for Performance Measurement, IEEE Access, 12: 33949-33958.

[11] Tainiyat Hanchinal, K. and Vaishali Bhavani, D. 2024. Clothes Price Comparison Using Machine Learning, International Research Journal of Engineering and Technology.