International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

CAR SALES FORECASTING USING PRICE AND FUEL TYPE ANALYSIS & PREDICTION

[1] Y Ramya, [2] M Lasya Priya, [3] G Sravan Kumar, [4] J Akshay
[1]AssistantProfessor, [234]Students
Department of Data Science
Siddhartha Institute of Technology & Sciences, Narapally
sits.ramyayelapati@siddhartha.co.in, 23TQ1A6750@siddhartha.co.in, 23TQ1A6745@siddhartha.co.in, 23TQ1A6713@siddhartha.co.in,

Abstract

The Car Sales Forecasting Using Price and Fuel Type Analysis and Prediction system is a data-driven application designed to analyze historical car sales information and predict future sales patterns. The automobile market contains large amounts of information related to car prices, fuel types, brands, models, sales quantities, and customer preferences. Analyzing these factors can help understand market trends and support better planning and decision-making.

The proposed system collects historical car sales data containing attributes such as car model, manufacturer, selling price, fuel type, transmission type, year, location, and sales volume. Data preprocessing techniques are applied to clean the dataset, handle missing values, remove duplicate records, and standardize numerical and categorical attributes. The processed data is then used for exploratory analysis and prediction.

The system analyzes the relationship between car prices, fuel types, and sales performance. It can compare the sales of petrol, diesel, electric, hybrid, and other fuel categories and examine how price variations influence customer demand. Data visualization techniques such as bar charts, line graphs, scatter plots, and comparison charts are used to present these patterns clearly.

A forecasting module uses statistical or Machine Learning techniques to estimate future car sales based on historical patterns and selected input features. The prediction model can be trained using historical sales records and evaluated using suitable performance metrics. Users can enter relevant car attributes and obtain an estimated sales prediction from the trained model.

Overall, the proposed system combines data analysis, visualization, and Machine Learning-based forecasting to provide useful insights into the automobile sales market. It can support dealers, manufacturers, analysts, and business managers in understanding price and fuel-type trends. Future enhancements can include real-time market data, advanced forecasting models, customer segmentation, regional demand prediction, electric-vehicle trend analysis, and interactive business intelligence dashboards.

I. Introduction

The automobile industry is a highly competitive market where customer preferences, vehicle prices, fuel types, economic conditions, and technological developments influence sales. Car manufacturers and dealers need to understand these factors to plan production, pricing, inventory, and marketing strategies. Historical sales data provides valuable information that can be analyzed to identify important market patterns.

Car price is one of the major factors that can influence customer purchasing decisions. Different price ranges attract different customer segments, and changes in vehicle prices can affect sales volume. Analyzing the relationship between price and sales can help identify price ranges associated with different levels of demand.

Fuel type is another important factor in the automobile market. Petrol and diesel vehicles have traditionally represented a significant portion of car sales, while electric and hybrid vehicles have gained increasing attention. Studying sales according to fuel type can help identify changes in consumer preferences and the development of different vehicle categories.

Traditional sales analysis often relies on spreadsheets and manually prepared reports. While these methods can provide basic information, they may not be efficient when working with large datasets or multiple variables. Data analytics and visualization can provide a more systematic approach to understanding sales trends, price distributions, and fuel-type preferences.

Machine Learning provides additional capabilities by using historical data to build predictive models. A forecasting model can learn relationships between car attributes and sales patterns and use those relationships to estimate future sales. The proposed system combines descriptive analysis, price and fuel-type analysis, visualization, and predictive modeling to provide a comprehensive approach to car sales forecasting.

II. Literature Survey

Data analytics has become an important technique in the automobile industry for understanding vehicle sales and customer preferences. Historical sales datasets can contain information about vehicle models, prices, fuel types, manufacturing years, locations, and sales quantities. Researchers and businesses use these datasets to identify trends and relationships between different vehicle attributes and sales performance.

Price analysis is commonly used to understand the relationship between product pricing and customer demand. In the automobile market, vehicle price can vary significantly based on brand, model, engine type, features, age, and fuel type. Statistical analysis and visualization techniques can be used to examine price distributions and identify patterns between vehicle prices and sales.

Fuel-type analysis has become increasingly relevant because the automobile industry is experiencing changes in consumer preferences and vehicle technology. Studies comparing petrol, diesel, hybrid, and electric vehicles can provide information about the adoption of different fuel categories. Historical datasets can be analyzed to determine how sales distributions vary between these categories.

Data preprocessing is an important stage in car sales analytics because real-world datasets may contain missing values, duplicate records, inconsistent price formats, incorrect entries, and categorical variations. Cleaning and transforming the data before analysis helps improve the reliability of results. Feature engineering can also be used to derive useful attributes such as vehicle age, price range, or sales period.

Machine Learning techniques have been applied to sales forecasting and demand prediction in various industries. Regression algorithms can predict numerical sales quantities, while time-series models can identify trends and seasonal patterns in historical sales. Ensemble learning methods can also be used when multiple vehicle attributes influence the prediction. Model performance should be evaluated using appropriate metrics and validated against unseen data.

Recent analytical systems increasingly combine exploratory data analysis, visualization, and predictive modeling within interactive dashboards. Such platforms allow users to examine historical trends and obtain forecasts through a single interface. The proposed Car Sales Forecasting system follows this approach by analyzing price and fuel-type relationships and using historical sales information to generate data-driven predictions.

III. System Analysis

The Car Sales Forecasting system is designed to analyze historical automobile sales data and predict future sales patterns using data analytics and Machine Learning techniques. The system accepts datasets containing information such as car model, manufacturer, price, fuel type, year, location, and sales volume. The input data is first validated to ensure that the required fields are available and correctly formatted. A preprocessing module removes duplicate records and handles missing, inconsistent, or invalid values. Numerical and categorical attributes are transformed into suitable formats for analysis and model training. Exploratory analysis is performed to understand sales distributions and relationships between price, fuel type, and sales volume. Visualization components present these relationships through charts, graphs, and comparison dashboards. The system identifies sales patterns across different price ranges and fuel categories. A predictive model is trained using historical sales data and selected vehicle attributes. The trained model is evaluated using appropriate performance measures to understand its prediction quality. Users can provide relevant car attributes to obtain an estimated sales prediction. The system therefore combines historical analysis and predictive modeling to support understanding of automobile sales trends.

Existing System

Existing car sales analysis is commonly performed using spreadsheets, database queries, and manually prepared business reports. Sales teams or analysts collect historical records and calculate sales totals, average prices, and category-wise performance manually. Basic reports may provide information about the number of cars sold but may not clearly show relationships between price, fuel type, and sales volume. Large datasets can make manual filtering and comparison time-consuming. Traditional systems may also require users to create individual charts for different fuel categories or price ranges. Historical sales trends may be available, but future sales prediction is often handled separately or not provided at all. Data inconsistencies and missing values can also affect manually generated analysis. Different sources may contain different formats for prices, vehicle models, and fuel categories. Identifying relationships between multiple vehicle attributes can be difficult using basic reports. Standard reporting systems may provide limited interactive visualization and forecasting capabilities. Users may therefore need additional analytical tools to perform predictive analysis. These limitations create a need for an integrated platform that combines sales analysis, price and fuel-type comparison, visualization, and forecasting.

Disadvantages of Existing System

- Manual sales analysis requires considerable time and effort.
- Large datasets are difficult to analyze using spreadsheets.
- Limited relationship analysis between price and sales.
- Fuel-type comparisons may require manual processing.
- Data cleaning may need significant manual effort.
- Static reports provide limited interactivity.
- Future sales forecasting may not be available.
- Difficult to identify complex sales patterns.
- Multiple reports may be required for different categories.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

- Limited predictive decision-support capabilities.
- Manual chart creation can be repetitive.
- Historical and predictive analysis may exist in separate systems.

Proposed System

The proposed Car Sales Forecasting system provides an integrated platform for analyzing car sales data and predicting future sales patterns. Users can upload historical car sales datasets containing attributes such as model, manufacturer, price, fuel type, year, location, and sales volume. The system validates the uploaded data and performs preprocessing to handle missing values, duplicate records, inconsistent formats, and invalid entries. Cleaned data is transformed into a suitable format for exploratory analysis and Machine Learning. The analytics module examines sales trends and relationships between vehicle price, fuel type, and sales volume. Users can compare petrol, diesel, electric, hybrid, and other fuel categories using interactive visualizations. Price-based analysis can identify sales patterns across different vehicle price ranges. A forecasting module trains a Machine Learning model using historical sales records and relevant vehicle attributes. The model can generate estimated sales values based on user-provided inputs and historical patterns. Model performance is evaluated using suitable metrics to provide information about prediction quality. Dashboards present historical trends, analytical comparisons, prediction results, and important KPIs in a clear format. The proposed system therefore provides a unified environment for automobile sales analysis and predictive forecasting.

Advantages of Proposed System

- Provides centralized car sales analysis.
- Analyzes the relationship between price and sales.
- Compares sales across different fuel types.
- Automates data preprocessing and analysis.
- Provides interactive charts and dashboards.
- Identifies historical sales trends.
- Supports Machine Learning-based sales prediction.
- Reduces manual report preparation.
- Allows users to analyze different price ranges.
- Helps identify changing fuel-type preferences.
- Provides data-driven forecasting insights.
- Can be extended with real-time automobile market data.

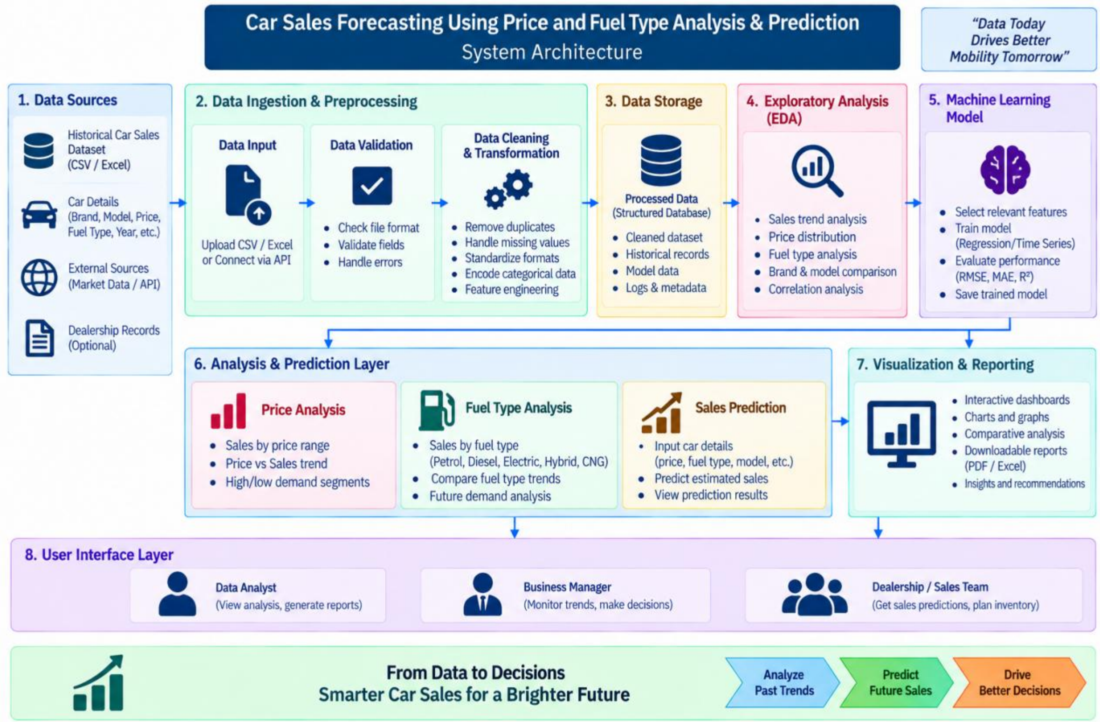
IV. Methodology

The methodology begins with collecting historical car sales data from an authorized dataset or business data source. The collected dataset is validated to check its structure, required fields, data types, and completeness. Data preprocessing is then performed to remove duplicate records and handle missing, inconsistent, or invalid values. Attributes such as car price, fuel type, manufacturing year, model, location, and sales volume are standardized for analysis. Exploratory Data Analysis is performed to understand sales distributions and relationships between important variables. Price analysis is used to examine sales performance across different price ranges and vehicle categories. Fuel-type analysis compares sales patterns among petrol, diesel, electric, hybrid, and other categories available in the dataset. Visualization techniques are used to represent trends and relationships through bar charts, line graphs, scatter plots, and interactive dashboards. Relevant features are selected and transformed for Machine Learning model development. A forecasting model is trained using historical

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

sales data and evaluated using appropriate prediction metrics. Users can enter vehicle-related attributes to generate an estimated sales prediction from the trained model. The complete methodology combines data collection, preprocessing, exploratory analysis, price analysis, fuel-type analysis, visualization, predictive modeling, evaluation, and reporting.

System Architecture



The system architecture consists of multiple interconnected layers that work together to analyze car sales data and generate predictions. The Data Input Layer receives historical car sales records through CSV, Excel, database, or authorized API sources. The Data Validation Layer checks file formats, required fields, data types, and basic data quality. The Data Preprocessing Layer handles missing values, duplicate records, inconsistent formats, and invalid entries before analysis. The Data Storage Layer stores cleaned historical sales data and model-related information securely. The Exploratory Analytics Layer analyzes sales volume, price distribution, vehicle categories, and historical trends. The Price and Fuel Analysis Layer examines relationships between car prices, fuel types, and sales performance. The Visualization Layer presents analytical results through charts, graphs, tables, and KPI cards. The Feature Engineering Layer prepares relevant attributes for predictive model training. The Machine Learning Layer trains and evaluates a forecasting model using historical sales data. The Prediction Layer accepts vehicle attributes and generates estimated sales values based on the trained model. The Reporting Layer provides analytical summaries, prediction results, and downloadable reports. Finally, the User Interface Layer presents dashboards, analysis results, and forecasts to authorized users in an easy-to-understand format.

V. Result and Output

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

V. Conclusion

The Car Sales Forecasting Using Price and Fuel Type Analysis and Prediction system provides a data-driven approach for understanding automobile sales patterns and estimating future sales. The system combines data preprocessing, exploratory data analysis, visualization, and Machine Learning techniques to transform historical car sales records into useful analytical information. This reduces the need for extensive manual analysis.

The system analyzes important factors such as car price, fuel type, vehicle model, manufacturer, year, location, and sales volume. Price-based analysis helps identify sales patterns across different price ranges, while fuel-type analysis provides insights into the distribution of petrol, diesel, electric, hybrid, and other vehicles. These analyses help users understand historical market behavior.

The forecasting module uses historical sales data and relevant vehicle attributes to generate estimated future sales values. The system also provides charts, graphs, KPI cards, and comparison dashboards to make the analytical and prediction results easier to understand. Model evaluation metrics can be used to assess the performance of the forecasting model before using its predictions.

The proposed system can be useful for automobile dealers, manufacturers, sales teams, analysts, and business managers for understanding sales trends and supporting inventory and planning activities. However, predictions are estimates based on historical data and should be considered as decision-support information rather than guaranteed future sales.

In the future, the system can be enhanced with real-time market data, advanced time-series forecasting, customer preference analysis, regional demand prediction, electric-vehicle sales forecasting, competitor-price analysis, and automated recommendation systems. These improvements can make the platform more comprehensive and useful for automobile sales analysis and forecasting.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.