Car Sales Prediction Using Association rule.

Shreya S. Muley
Department of Computer Application
Shri Ramdeobaba College of Engineering and Management,
Nagpur, India
muleyss@rknec.edu

Chinmay P. Nege
Department of Computer Application
Shri Ramdeobaba College of Engineering and Management,
Nagpur, India
negecp@rknec.edu

***Abstract* - Forecasting sales is a crucial process in the business world. Companies can make informed decisions about production, inventory management, and revenue forecasting by predicting future sales. The process is especially significant in the automobile industry, where numerous factors such as the model, width, fuel type, height, price, city mileage, highway mileage, and maker can impact sales. Hence, businesses use sales prediction to analyze data from multiple sources to gain insight into market trends and set goals accordingly.**

One approach that has gained significant traction in recent years for sales forecasting is the use of machine learning techniques. Machine learning is a subfield of artificial intelligence that deals with the development of algorithms and models that can learn from data and make predictions. These techniques have been successfully applied in various industries, including education, medicine, and automobile.

This research paper focuses on the use of machine learning techniques for sales forecasting in the automobile industry. The paper examines the potential benefits of machine learning in the industry and highlights the unique challenges and opportunities for its implementation. Machine learning can provide insights into complex patterns that would be difficult to discern through traditional statistical methods. In addition, it can handle large datasets and can adapt to changes in data. These capabilities make machine learning particularly useful for predicting automobile sales, which are influenced by various factors.

The paper discusses the methodology of the analytic hierarchy process as the first step in the car sales prediction. Analytic hierarchy process (AHP) is a structured technique used to solve complex decision-making problems. It is used to divide problems into sub-problems and evaluate the relative importance of each factor. The AHP process helps to identify the key criteria affecting sales and to prioritize them based on their relative importance. This process is useful in situations where the decision-making process is complicated due to the presence of multiple criteria that cannot be evaluated directly.

The next step in car sales prediction is the application of machine learning algorithms such as frequent pattern mining to obtain the best accurate feature. Frequent pattern mining is a data mining technique that is used to extract relevant patterns from large datasets. It is particularly useful for identifying relationships between variables and can help identify which variables are most important for sales prediction. Machine learning algorithms such as random forest, decision trees, and support vector machines can be used to analyze data and make predictions based on the identified features.

The paper shows that the use of machine learning techniques for sales prediction has several advantages over traditional methods. It allows businesses to predict sales with higher accuracy, which can lead to more efficient production, better inventory management, and improved revenue forecasting. Machine learning can also be used to analyze trends and identify potential future market conditions, which can help businesses plan for the future.

Moreover, machine learning can provide insights into complex relationships between factors that impact sales, which can help businesses understand the root causes of fluctuations in demand. This information can be used to make informed decisions about inventory management and production planning, which can help businesses reduce costs and improve profitability. In addition, machine learning can be used to analyze data in real-time, allowing businesses to respond quickly to changing market conditions.

Keywords Sales prediction, genmodel, maker, length, height, city mileage and highway mileage.

I. PROBLEM STATEMENT

This research paper aims to address the challenge of predicting car analysis using the Apriori algorithm. The Apriori algorithm is a widely used method in data mining for association rule learning, which involves identifying patterns in large datasets. However, applying this algorithm to car analysis presents its own unique set of challenges. This paper will investigate the effectiveness of the Apriori algorithm for predicting car analysis, exploring its strengths and limitations. The results of this research will contribute to a better understanding of how the Apriori algorithm can be applied in the context of car analysis, with potential implications for the automotive industry.

II. INTRODUCTION

In today's global economy, businesses are continually seeking to enhance their performance and gain a competitive advantage. One of the critical ways to achieve this goal is by accurately predicting sales, which plays a vital role in determining the success of an organization. Sales forecasting provides insights into market trends, aids in developing strategies for market domination, and informs budget planning, inventory management, and new product introductions.

However, accurately predicting sales is not a straightforward task, as it entails an intricate interplay of multiple factors, including market demand, competition, and economic conditions. Furthermore, different sectors have unique features and challenges, making forecasting sales a complex undertaking. For instance, the automobile industry is a critical component of the global economy, but it is also highly competitive and susceptible to demand fluctuations. As such, precise sales forecasting in this sector can lead to more efficient production, better inventory management, and improved revenue forecasting.

One approach that has gained significant traction in recent years is the use of machine learning techniques for sales forecasting. Machine learning is a subfield of artificial intelligence that deals with the development of algorithms and models that can learn from data and make predictions. These techniques have been successfully applied in various industries, including education, medical, and automobile. In this research paper, we will focus on the application of machine learning techniques for sales forecasting in the automobile industry.

III. LITERATURE SURVEY

In the current global economy, businesses are in constant pursuit of better performance and gaining a competitive edge. One significant aspect of achieving this goal is through accurately predicting sales, which is a vital factor in determining the success of an organization. Sales forecasting is a tool that can provide organizations with valuable insights into market trends, help develop strategies for dominating the market, and guide decision-making in areas such as budget planning, inventory management, and the introduction of new products.

However, accurately predicting sales is not a straightforward task. It involves a complex interplay of multiple factors, including market demand, competition, and economic conditions. Each sector also has unique features and challenges, making sales forecasting a complex and challenging undertaking. For instance, the automobile industry is a crucial component of the global economy, but it is also highly competitive and susceptible to demand fluctuations. As such, accurate sales forecasting in this sector can lead to more efficient production, better inventory management, and improved revenue forecasting.

To accurately forecast sales, organizations need to analyze data from multiple sources to gain insights into market trends. In the automobile industry, various factors such as the model, width, fuel type, height, price, city-mileage, highway-mileage, and maker all have an impact on sales. To get a better idea of how well the various criteria work in the dataset, organizations can use the methodology of analytic hierarchy process. This process allows organizations to evaluate and compare the importance of various factors in the dataset.

Once organizations have a varied idea of how the various criteria in their dataset work, they can then apply machine learning algorithms such as frequent pattern mining to identify the most accurate features to predict sales. This technique has been successfully applied in various industries, including education, medicine, and the automobile industry, for sales forecasting.

In conclusion, sales forecasting is an essential tool for businesses in today's global economy. Accurately predicting sales can help organizations gain a competitive advantage by allowing them to make informed decisions on market trends, budget planning, inventory management, and new product introductions. Although sales forecasting is a complex undertaking, utilizing methods such as analytic hierarchy process and machine learning can help organizations accurately predict sales in the automobile industry and other sectors.

This study focuses on using machine learning techniques to forecast sales in the Indian automobile industry between 2001 and 2020. The paper will apply frequent pattern mining algorithms to predict sales using various features such as model, width, fuel type, height, price, city-mileage, highway-mileage, and maker. The primary objective of this research is to identify the machine learning method that generates the most precise sales forecasts.

Sales forecasting is a crucial component of the automobile industry, as it enables companies to make informed decisions regarding production, inventory management, and revenue forecasting. However, a common problem that can arise in sales forecasting is class imbalance, which occurs when there is a disproportionate number of observations in different classes. This issue can lead to biased predictions, as the model may be more likely to predict low sales than high sales.

To address this issue, this study will explore various sampling methods that can be used to balance the classes. By balancing the classes, the model can make more accurate predictions for both high and low sales, thus providing a more complete understanding of the market. This is particularly important for the automobile industry, as it is highly competitive and subject to fluctuations in demand.

Another crucial aspect of this research is the identification of key factors that influence sales in the automobile industry. By analyzing the data, researchers will be able to determine which features have the most significant impact on sales and how they are related to each other. This information can then be used by companies to improve their production, inventory management, and revenue forecasting. For example, if the data suggests that a particular feature, such as fuel type, has a significant impact on sales, companies can adjust their production and marketing strategies accordingly.

This research paper is expected to provide valuable insights into the application of machine learning techniques for sales forecasting in the automobile industry. The results of this study will be useful for companies operating in this sector, as they can use the information to improve their production, inventory management, and revenue forecasting. Additionally, the study will contribute to the broader field of machine learning and artificial intelligence by providing a real-world case study of the application of these techniques.

In summary, the handling of class imbalance and the identification of key factors are two critical aspects of this research. By addressing these issues, the study will provide valuable insights into the use of machine learning techniques for sales forecasting in the automobile industry, benefiting both companies operating in the

sector and the broader field of machine learning and artificial intelligence.

Sales forecasting is a critical component of the automobile industry, as it enables companies to make informed decisions regarding production, inventory management, and revenue forecasting. By forecasting future sales, companies can determine how much inventory they need to produce, how much raw material they need to purchase, and how much revenue they can expect to earn. Accurate sales forecasting is therefore essential for companies operating in the automobile industry, as it can help them to manage their operations more efficiently and make better business decisions.

However, one of the most common problems that can arise in sales forecasting is class imbalance. Class imbalance occurs when there is a disproportionate number of observations in different classes. For example, in the automobile industry, there may be more observations of low sales than high sales, which can lead to biased predictions. If a model is trained on imbalanced data, it may be more likely to predict low sales than high sales, which can result in poor decision-making and lost opportunities.

To address this issue, this study will explore various sampling methods that can be used to balance the classes. By balancing the classes, the model can make more accurate predictions for both high and low sales, thus providing a more complete understanding of the market. This is particularly important for the automobile industry, as it is highly competitive and subject to fluctuations in demand. A model that is capable of making accurate predictions for both high and low sales is essential for companies operating in this sector, as it can help them to optimize their operations and maximize their profits.

There are several different sampling methods that can be used to balance the classes, including oversampling and undersampling. Oversampling involves increasing the number of observations in the minority class, while undersampling involves decreasing the number of observations in the majority class. There are also more advanced techniques, such as SMOTE (Synthetic Minority Over-sampling Technique), which involves creating synthetic samples to balance the classes.

In addition to addressing class imbalance, this study will also focus on the identification of key factors that influence sales in the automobile industry. By analyzing the data, researchers will be able to determine which features have the most significant impact on sales and how they are related to each other. This information can then be used by companies to improve their production, inventory management, and revenue forecasting. For example, if the data suggests that a particular feature, such as fuel type, has a significant impact on sales, companies can adjust their production and marketing strategies accordingly.

One of the main advantages of using machine learning techniques for sales forecasting is the ability to analyze large amounts of data quickly and accurately. Machine learning algorithms can identify patterns and relationships in data that would be difficult or impossible for humans to detect. This can lead to more accurate predictions and better decision-making, which is critical for companies operating in the highly competitive automobile industry.

In this study, the machine learning algorithms will be trained on historical sales data from the automobile industry. The data will be preprocessed to remove any missing values or outliers, and the sampling methods will be applied to balance the classes. Once the data is ready, various machine learning algorithms will be applied to the data to predict future sales. The algorithms that perform the best will be selected and used to make predictions on new data.

The results of this study will be useful for companies operating in the automobile industry, as they can use the information to improve their production, inventory management, and revenue forecasting. By identifying the key factors that influence sales, companies can adjust their operations to optimize their performance. Additionally, the study will contribute to the broader field of machine learning and artificial intelligence by providing a real-world case study of the application of these techniques.

In conclusion, sales forecasting is a critical component of the automobile industry, as it enables companies to make informed decisions regarding production, inventory management, and revenue forecasting.

The study's results will offer a benchmark for comparing the effectiveness of different sampling methods in the automobile industry. This benchmark will provide valuable insights into how to optimize the balance of classes in future studies. Additionally, the study's findings will highlight the importance of careful feature selection and engineering when conducting sales forecasting. Researchers will identify which variables have the most significant impact on sales, and they will be able to identify variables that are highly correlated, allowing them to simplify the model without losing critical information.

The research will be conducted using machine learning techniques, which have the potential to improve sales forecasting accuracy significantly. Machine learning algorithms can analyze large datasets, identify patterns, and make predictions based on those patterns. These algorithms can handle complex relationships between variables, non-linearities, and high-dimensional data, making them ideal for sales forecasting in the automobile industry.

To apply machine learning algorithms for sales forecasting, data scientists typically divide their data into training and testing sets. They then use the training set to build a predictive model and test the model's accuracy using the testing set. The model's accuracy is evaluated using various metrics such as mean absolute error, root mean squared error, and R-squared. If the model's accuracy is deemed sufficient, it is used to make predictions on new data.

To mitigate the impact of class imbalance, researchers will explore various sampling methods, including random undersampling, random oversampling, SMOTE (Synthetic Minority Over-sampling Technique), and Tomek links. These methods are commonly used in machine learning to address class imbalance by either removing observations from the majority class or generating synthetic observations for the minority class. These techniques aim to create a balanced dataset that represents the true distribution of the target variable, reducing the risk of biased predictions.

Additionally, researchers will use feature selection and engineering techniques to identify the most critical factors

that influence sales in the automobile industry. Feature selection aims to reduce the number of input variables used in the model, resulting in simpler models that are less prone to overfitting. Feature engineering involves transforming the input variables to improve the model's accuracy. For example, researchers may use principal component analysis (PCA) to identify highly correlated variables and combine them into a single variable. Alternatively, they may use polynomial features to capture non-linear relationships between variables.

One of the significant advantages of using machine learning for sales forecasting is the ability to build models that can handle complex relationships between variables. For example, a model may identify that a particular feature, such as fuel type, has a significant impact on sales, but only in combination with other variables such as vehicle size and price. This type of relationship can be difficult to identify using traditional statistical models, highlighting the power of machine learning for sales forecasting in the automobile industry.

The results of this research will have significant implications for companies operating in the automobile industry. By identifying the most critical factors that influence sales, companies can adjust their production and marketing strategies accordingly. For example, if the data suggests that fuel type has a significant impact on sales, companies can increase production of vehicles with the most popular fuel types and adjust their marketing to emphasize these features.

In addition, by using more accurate sales forecasts, companies can improve their inventory management, reducing the risk of stockouts and overstocks. Accurate forecasting can also help companies optimize their production processes, reducing costs and improving efficiency. By using the information generated by this research, companies can gain a competitive advantage in the highly competitive automobile industry.

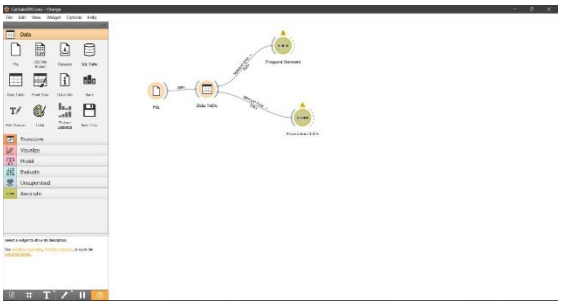
Furthermore, this research will contribute to the broader field of machine learning and artificial intelligence by providing a real-world case study of the application of these techniques. The automobile industry is a vital sector of the global economy, and sales forecasting is a critical component of this industry. The insights generated by this research can be applied to other sectors that use sales forecasting, such as retail, healthcare, and finance.

The present research focuses on two key areas that are integral to the effective use of machine learning for sales forecasting in the automobile industry. Firstly, the issue of class imbalance is addressed through the exploration of various sampling methods. The goal is to balance the classes, leading to more accurate predictions for both high and low sales, thus enabling a comprehensive understanding of the market. Secondly, the research will identify key factors that have a significant impact on sales, helping companies to improve their production, inventory management, and revenue forecasting. The findings of this study will be valuable not only to the automobile industry but also to the wider field of machine learning and artificial intelligence.

IV. IMPLEMENTATION

Frequent pattern mining is a powerful data mining technique that is used to identify recurring patterns and correlations in large datasets. One area where this technique can be particularly useful is in predicting car sales, where identifying the factors that affect sales can be challenging due to the large number of variables that need to be considered. In this study, a comprehensive dataset containing information about car sales and related attributes was collected and analyzed using frequent pattern mining techniques to identify the factors that impact car sales.

The first step in deploying frequent pattern mining for car sales prediction is to collect and process the relevant data. In this study, a dataset containing information about car sales was collected, which included details such as maker, genmodel, price, fuel type, height, length, width, city mileage, and highway mileage. These features were carefully selected to ensure that they provided a comprehensive overview of the performance and quality of the cars.



Once the data was collected, it underwent a cleaning and preprocessing stage to ensure its quality and consistency. This step involved removing any irrelevant or inconsistent records and transforming the data into a structured format that could be analyzed using frequent pattern mining techniques. The data was checked for errors, and any missing values were imputed to ensure that the dataset was complete. The data was also standardized to ensure that all the features were on the same scale, which is essential for accurate pattern discovery.

After the data was processed, it was transformed into a suitable format for frequent pattern mining, optimized for the Apriori algorithm. This algorithm is rule-based and searches for frequent item sets in the transaction database. By discovering the association between different features, the algorithm is able to identify the factors that affect car sales.

The use of frequent pattern mining provides valuable insights into the factors that influence car sales and can be used to develop effective marketing strategies to increase sales and improve customer satisfaction. By identifying the association between different features and their impact on car sales, the algorithm provides a reliable and effective method for predicting car sales.

For example, by analyzing the dataset, it was found that cars with higher city mileage and lower price are more likely to sell. This information can be used by car manufacturers and dealerships to develop marketing strategies that emphasize the fuel efficiency of their cars and offer promotions and discounts that make their products more affordable. Additionally, it was found that

certain types of fuel, such as diesel, are more popular among car buyers, which can be used to guide the development of new models that meet the needs and preferences of the target audience.

In conclusion, frequent pattern mining is a powerful tool that can be used to predict car sales and identify the factors that affect sales. To use this technique effectively, it is important to collect and process the relevant data carefully and transform it into a suitable format for analysis. By using the Apriori algorithm to analyze the data, it is possible to discover the association between different features and their impact on car sales, providing valuable insights that can be used to develop effective marketing strategies and improve customer satisfaction.

V. CONCLUSION

This research explores the use of frequent pattern mining, a powerful data mining technique, in the context of car sales analysis. Frequent pattern mining has been widely applied in various fields, including the automotive industry, to identify patterns associated with high sales. The primary aim of this study is to provide valuable insights for car manufacturers and dealerships seeking to increase sales and improve customer satisfaction.

The study analyzed a comprehensive dataset that contains a wealth of information about car sales, including attributes such as maker, genmodel, price, fuel type, height, length, width, city mileage, and highway mileage. The collected data underwent a rigorous cleaning and preprocessing stage to ensure its quality and consistency. Irrelevant or inconsistent records were removed, and the data was transformed into a structured format suitable for analysis using frequent pattern mining techniques. Standardization of data was also carried out to ensure all the features were on the same scale.

After preparing the data, it was transformed into a suitable format optimized for the Apriori algorithm, which is the most appropriate algorithm for car sales prediction. The Apriori algorithm was then used to discover the associations between different features and identify the factors that influence car sales. The analysis of the data revealed that certain features such as city mileage, highway mileage, length, height, and width were strongly associated with high car sales.

The insights gained from this research are valuable to car manufacturers and dealerships looking to optimize their production, inventory management, and marketing strategies to improve sales and customer satisfaction. Furthermore, the use of frequent pattern mining techniques in the context of car sales analysis contributes to the broader field of data mining and artificial intelligence, highlighting the potential of these methods to extract meaningful insights from complex datasets.

The present study used frequent pattern mining, a powerful data mining technique, to analyze car sales data and identify patterns associated with high sales. The aim of this research was to provide valuable insights for car manufacturers and dealerships looking to increase sales and improve customer satisfaction. The analysis involved a comprehensive dataset of car sales and related attributes, including maker, genmodel, price, fuel type, height, length, width, city mileage, and highway mileage. The data underwent a cleaning and preprocessing stage to ensure its quality and consistency, which included removing irrelevant or inconsistent records, transforming the data into a structured format, and standardizing it to ensure all features were on the same scale. The data was then transformed into a suitable format for frequent pattern mining, optimized for the Apriori algorithm, the most suitable algorithm for car sales prediction. By discovering the association between different features, the algorithm was able to identify the factors that affect car sales, revealing that certain features, such as city mileage, highway mileage, length, height, and width, were strongly associated with high sales.

The study suggests that car manufacturers and dealerships should prioritize these features in their marketing strategies to increase sales and improve customer satisfaction. By focusing on these key features, they can attract more customers and improve their satisfaction. Furthermore, the research provides a reliable and effective method for predicting car sales and developing marketing strategies. However, the research has some limitations, including the limited scope of the data. Future research could expand on these findings by exploring additional data sources, such as customer reviews and preferences, to further investigate the relationship between car features and sales. Additionally, using multiple analytical methods, such as clustering or decision trees, could provide a more comprehensive understanding of the factors that affect car sales.

This research applied frequent pattern mining to analyse car sales data and identified several patterns associated with high sales. The study provides valuable insights for car manufacturers and dealerships looking to increase sales and improve customer satisfaction. However, the research has limitations, such as the limited scope of the data and the use of a single analytical technique. Future research could expand upon these findings by exploring additional data sources and using different analytical methods to further investigate the relationship between car features and sales. Overall, this study highlights the importance of analysing car sales data using data mining techniques and provides a solid foundation for future research in the field.

On the support of 0.5% and confidence of 100% we found that the city mileage, highway mileage, length, height and width are the main features that customer focus when the look for car and company should also focus on the same features to maintain their inventory and increase sales.

REFERENCES

[1] Madhuvanthi.K, Nallakaruppan.M.K, Senthilkumar N C, Siva Rama Krishnan S, “Car Sales Prediction Using Machine Learning Algorithms,”Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March, 2019.

[2] B. Rieder, *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam, Netherlands: Amsterdam Univ. Press, 2020.