**Data Mining DS 600 Capstone Project**

**Consumer Behaviour Analysis for Electric Vehicle Adoption using Digital Twins**

**Problem Statement:**

As the world transitions towards sustainable transportation solutions, electric vehicles (EVs) have emerged as a promising alternative to traditional gasoline-powered vehicles. However, the widespread adoption of EVs is still hindered by various factors, including high prices, limited range, and lack of charging infrastructure. This study aims to analyse a comprehensive dataset containing information about electric cars to gain valuable insights into the current state of the electric car market and identify potential areas for improvement.

**Dataset:**

The dataset encompasses a wide range of features, including 'Postal Code', 'Model Year', 'Make', 'Model', 'Electric Range', and 'Base MSRP'. Using these features, the primary objective is to develop predictive models capable of accurately forecasting the 'Electric Vehicle Type' that consumers are most likely to choose when making a purchase decision.

**Exploratory Data Analysis (EDA)**

Exploratory Data Analysis: Insights from Electric Vehicle Dataset

**Top Vehicle Makes:**

- Tesla dominates the market, accounting for a significantly larger share than other manufacturers.

- Nissan, Chevrolet, Ford, and BMW are among the other popular makes for electric vehicles in the dataset.

- This information can guide targeted marketing strategies and infrastructure development efforts based on consumer preferences.

**Electric Range by Model Year:**

- Newer model years generally have higher electric ranges, with some outliers offering exceptionally long ranges.

- The electric range has steadily improved over the years, addressing one of the key barriers to widespread electric vehicle adoption.

- This trend highlights the rapid technological advancements in the electric vehicle industry and can inform consumer decision-making processes.

**Correlation Analysis:**

- The 'Postal Code' feature shows no correlation with other numerical features, suggesting that the geographic distribution of electric vehicles may be influenced by factors not captured in the dataset.

- 'Model Year' has a moderate negative correlation (-0.48) with 'Electric Range', indicating that older model years tend to have lower electric ranges.

- 'Base MSRP' has a weak positive correlation (0.11) with 'Electric Range', suggesting that higher-priced electric vehicles generally offer longer ranges.

**Popular Electric Vehicle Makes and Models:**

- The Tesla Model 3 and Tesla Model Y are the most prevalent models, aligning with the dominance of Tesla in the overall market share.

- Other popular models include the Nissan LEAF, Chevrolet Bolt EV, Ford Mustang Mach-E, and Volkswagen ID.4.

- This information can help advocacy groups tailor their marketing efforts and infrastructure planning based on the most popular models in the region.

**Predictive Modeling**

Building upon the insights gained from the EDA, we implemented various machine learning models to predict the 'Electric Vehicle Type' that consumers are most likely to choose based on the provided features. The following models were evaluated and the scores of each model are displayed below:

**1. Naive Bayes**

- Accuracy: 0.8964373395030351

- ROC AUC Score: 0.8193302711926629

- Precision: 0.92

- Recall: 0.96

- F1-Score: 0.94

**2. Logistic Regression**

- Accuracy: 0.8344234337722521

- ROC AUC Score: 0.6884668793554423

- Precision: 0.86

- Recall: 0.95

- F1-Score: 0.90

**3. Random Forest Classifier**

- Accuracy: 0.9999062950954853

- ROC AUC Score: 0.9999815568998911

- Precision: 1.00

- Recall: 1.00

- F1-Score: 1.00

The Random Forest Classifier and Naive Bayes models emerged as the top performers, with R-squared scores exceeding 0.90, indicating strong predictive capabilities. These models can be leveraged to guide future EV development and infrastructure planning efforts, ensuring a more informed and data-driven approach to supporting the state's transition to sustainable transportation.

**Recommendations and Conclusion**

Based on the analysis and findings, we recommend the following strategies for electric car advocacy groups to promote the widespread adoption of electric cars:

1. Urban Focus: Concentrate efforts on promoting the use of electric cars in urban areas, where the concentration of EVs is higher, and the impact on reducing emissions and congestion can be more significant.

2. Collaboration with Electric Utility Companies: Partner with electric utility companies to ensure the necessary charging infrastructure is in place and promote the use of renewable energy sources to power electric cars, aligning with sustainability goals.

3. Public Charging Station Advocacy: Advocate for the installation of more public charging stations, addressing range anxiety and increasing the convenience of owning an electric vehicle.

4. Targeted Marketing and Incentive Promotion: Utilize the insights from the make, model, and pricing analysis to develop targeted marketing campaigns and promote awareness of clean alternative fuel vehicle incentives in regions with lower adoption rates.

By leveraging the predictive models developed in this study and implementing the recommended strategies, electric car advocacy groups can effectively support the transition to a more sustainable transportation system. The widespread adoption of electric cars will not only reduce carbon emissions but also contribute to a cleaner and more environmentally friendly future.