**Assignment 6**

**Aim:**

To implement the Naive Bayes algorithm on a car dataset to predict the price category of used cars.

**Objective:**

Utilize the Naive Bayes classification algorithm to categorize used car prices based on attributes such as manufacturing year, kilometers driven, fuel type, seller type, transmission, and ownership history.

**Introduction**

Naive Bayes is a probabilistic classification method grounded in Bayes' Theorem. It operates under the assumption that features are conditionally independent given the class label. This assumption simplifies computation and enables efficient handling of high-dimensional data. Despite its simplicity, Naive Bayes often delivers performance comparable to more complex models.

**Why Use Naive Bayes for the Car Dataset:**

Naive Bayes is a suitable choice for this dataset due to the following reasons:

* **Handles Mixed Data Types:** Supports both categorical and numerical attributes.
* **Efficient with Small Datasets:** Performs well even with limited data.
* **Fast and Simple:** Quick training and easy to interpret results.
* **Probabilistic Output:** Offers insights into prediction confidence through probability scores.

**Advantages of Naive Bayes Algorithm:**

* **Simplicity:** Easy to understand and implement.
* **Scalability:** Efficient on larger datasets.
* **Compatibility with Categorical Data:** Works naturally with label-encoded variables.
* **Effective on Limited Data:** Shows good performance on small datasets.
* **Probability-based Results:** Provides class probabilities for informed decision-making.

**Dataset Description**

* **Filename:** car\_data.csv
* **Attributes:**
  + name – Car model
  + year – Year of manufacture
  + selling\_price – Selling price of the car
  + km\_driven – Total kilometers driven
  + fuel – Fuel type (Petrol/Diesel/CNG/etc.)
  + seller\_type – Whether sold by an individual or dealer
  + transmission – Manual or Automatic transmission
  + owner – Ownership history (e.g., First Owner)

**Implementation Steps:**

1. **Importing Libraries:**
   * Use pandas for data manipulation.
   * Use sklearn for preprocessing, modeling, and evaluation.
   * Use matplotlib and seaborn for data visualization.
2. **Loading and Inspecting the Dataset:**
   * Load the dataset into a Pandas DataFrame.
   * Check for missing values and data types.
3. **Data Preprocessing:**
   * Remove irrelevant features (e.g., name).
   * Encode categorical variables (fuel, seller\_type, transmission, owner).
   * Convert selling\_price into categorized bins such as "Low", "Medium", and "High".
4. **Feature Encoding:**
   * Apply LabelEncoder to transform categorical columns into numeric values.
5. **Dataset Splitting:**
   * Split the dataset into training and testing sets using train\_test\_split.
6. **Model Training:**
   * Train a GaussianNB classifier from scikit-learn, assuming numerical features follow a Gaussian distribution.
7. **Making Predictions:**
   * Predict the price category for test data samples.
8. **Model Evaluation:**
   * Evaluate model performance using metrics like:  
       
     + Accuracy
     + Classification Report
     + Confusion Matrix

**Conclusion**

The Naive Bayes classifier was successfully applied to predict car price categories using a combination of numerical and categorical data. The model offers not only accuracy but also interpretability, thanks to its probabilistic nature and clear evaluation metrics. This experiment highlights the practical use of probability-based classifiers in understanding and predicting data-driven trends.

**References:**

* [scikit-learn Naive Bayes Documentation](https://scikit-learn.org/stable/modules/naive_bayes.html)