# **ABC-XYZ Inventory Classification**

#### 

#### **Project Background and Objective**

The project aimed to develop a machine learning-based system for classifying inventory items using ABC analysis (based on sales value) and XYZ analysis (based on demand variability). The primary objective was to optimize stock management and enhance decision-making for businesses by prioritizing high-value and high-variability items. This classification helps in efficient inventory control, reducing costs, and improving forecasting accuracy, which is critical for supply chain management.

#### **CRISP-DM Framework Application**

##### **1. Business Understanding**

* **Focus**: The project addressed the business need for better inventory classification to prioritise stock management. ABC analysis categorises items into A (top 10% by sales value), B (next 20%), and C (remaining 70%), while XYZ analysis classifies based on demand variability (X for low, Y for medium, Z for high).
* **Objective**: To integrate these classifications with machine learning to automate and enhance decision-making, enabling businesses to focus on critical inventory items.

##### **2. Data Understanding**

* **Dataset Description**: The dataset comprised 1,000 inventory items with 18 columns, including:
  + **Item\_ID**: Unique identifier for each item.
  + **Item\_Name**: Name of the item.
  + **Category**: Item category (e.g., Grocery, Apparel).
  + **Jan\_Demand to Dec\_Demand**: Monthly sales data (units sold per month).
  + **Total\_Annual\_Units**: Total units sold in the year.
  + **Price\_Per\_Unit**: Selling price per unit.
  + **Total\_Sales\_Value**: Calculated as Total\_Annual\_Units × Price\_Per\_Unit.
* **Data Exploration**: Initial analysis involved checking for missing values and duplicates, confirming the dataset was clean. Visualizations, such as kernel density estimation (KDE) plots and correlation heatmaps, were used to understand data distributions and patterns. Notably, monthly demand figures showed high correlation, indicating consistent demand patterns, while price per unit had low correlation with demand, suggesting price did not significantly influence sales volume.

##### **3. Data Preparation**

* **Data Cleaning**: Ensured no missing or duplicate values, maintaining data quality for modeling.
* **Feature Engineering**:
  + **ABC Classification**: Used pandas.qcut to categorize items into A, B, C based on sales value quantiles, aligning with business rules (e.g., top 10% as A).
  + **XYZ Classification**: Calculated demand variability using the Coefficient of Variation (CV), categorizing items into X (low variability), Y (medium), and Z (high).
  + **Combined Feature**: Created a new feature, ABC\_XYZ class, by combining ABC and XYZ classifications for a holistic inventory categorization.
* **Preprocessing**: Standardized numerical features using StandardScaler to ensure normal distribution, facilitating machine learning model performance.
* **Train-Test Split**: Divided the dataset into 80% training and 20% testing sets to evaluate model generalization.

##### **4. Modeling**

* **Model Selection**: Implemented and compared multiple classification algorithms to predict the combined ABC\_XYZ classes, including:
  + DecisionTreeClassifier
  + RandomForestClassifier
  + KNeighborsClassifier
  + LogisticRegression
  + GradientBoostingClassifier
  + Support Vector Machine (SVM)
  + Naive Bayes
* **Training**: Each model was trained on the training dataset, leveraging Scikit-learn for implementation. The focus was on capturing the relationships between features (e.g., monthly demand, sales value) and the classification labels.

##### **5. Evaluation**

* **Metrics Used**: Evaluated models using a comprehensive set of performance metrics:
  + Accuracy: Overall correctness of predictions.
  + Precision: Proportion of positive identifications that were actually correct.
  + Recall: Proportion of actual positives correctly identified.
  + F1-score: Harmonic mean of Precision and Recall, balancing both metrics.
  + Confusion Matrix: Visualized to assess model performance across classes.
* **Results**:
  + Decision Tree Classifier achieved the highest accuracy of 98%, indicating excellent performance in classifying inventory items.
  + Random Forest and Gradient Boosting also performed well, closely following Decision Tree.
  + Naive Bayes was identified as the weakest performer, suggesting limitations in handling the dataset's complexity.
  + Visualizations, such as confusion matrices, highlighted Decision Tree's ability to correctly classify most items, particularly in high-value (A) and low-variability (X) categories.

#### **Technologies Employed**

The project leveraged the following tools and technologies, showcasing technical proficiency:

* **Programming Language**: Python 3, used for data analysis, modeling, and visualization.
* **Libraries/Frameworks**:
  + **Pandas**: For data manipulation and analysis, such as loading data and feature engineering.
  + **NumPy**: For numerical operations, supporting data preprocessing.
  + **Matplotlib and Seaborn**: For data visualization, including KDE plots, correlation heatmaps, and pie charts for category distribution.
  + **Scikit-learn**: For machine learning models, preprocessing (e.g., StandardScaler), and evaluation metrics.
* **Environment**: Google Colab, providing a cloud-based platform for running the Jupyter Notebook, facilitating collaborative development and execution.

#### **Results and Impact**

* The project successfully demonstrated the application of machine learning to inventory classification, achieving high accuracy with the Decision Tree model (98%). This outcome validated the approach of combining ABC and XYZ analyses with machine learning for enhanced inventory management.
* The results enabled businesses to prioritize inventory control, focusing on high-value and high-variability items, potentially reducing costs and improving forecasting accuracy. The use of visualizations and performance metrics ensured transparency and interpretability, critical for business adoption.

#### **Summary Table: Project Highlights**

| **Phase** | **Key Activities** | **Key Outcomes** |
| --- | --- | --- |
| Business Understanding | Identified need for inventory classification to optimize stock management. | Defined clear objectives for ABC-XYZ analysis. |
| Data Understanding | Analyzed dataset, visualized distributions, and checked data quality. | Confirmed clean dataset, identified demand patterns. |
| Data Preparation | Cleaned data, engineered ABC/XYZ features, standardized, and split into train/test. | Prepared high-quality data for modeling. |
| Modeling | Implemented Decision Tree, Random Forest, KNN, Logistic Regression, etc. | Selected models for classification. |
| Evaluation | Used Accuracy, Precision, Recall, F1-score, Confusion Matrix. | Decision Tree achieved 98% accuracy. |