

**The Superior University, Lahore**



|  |  |  |  |
| --- | --- | --- | --- |
| Course Title: | AI- Lab | | |
| Programme Name: | BSDS | | |
| Semester: | 3rd | Section: | BSDSM-3A |
| Student’s Name: | H.Hamza Dar | | |
| Instructor Name: | Prof. Rasikh Ali | | |

**Project Report**

**Store Type Classification Project Report**

**1. Introduction**

The goal of this project is to classify store types based on the available product data, including item names, prices, and related features. The classification is carried out using a Random Forest Classifier, which is evaluated using accuracy and classification metrics such as precision, recall, and F1-score.

**2. Dataset Overview**

The dataset contains information about different stores, the products they sell, and product prices. It includes the following columns:

* **Item Name**: The name of the item sold in the store.
* **Item Price**: The price of the item.
* **Store Type**: The type of store (target variable).

**3. Data Preprocessing**

Data preprocessing is a crucial step before building a machine learning model. The preprocessing steps included:

* **Mean Encoding**: The 'Item Name' feature was encoded using mean encoding, where each item name was replaced by the average price of the item.
* **Normalization**: The features 'Item Price' and 'TEncoded' (the encoded item prices) were normalized using MinMaxScaler to bring the values to the same scale (0 to 1).
* **Label Encoding**: The target variable 'Store Type' was label-encoded, converting categorical store types into numerical values.
* **One-Hot Encoding**: A one-hot encoding approach was used for the 'Item Name' feature to create binary columns representing each unique item.

**4. Model Development**

For the classification task, a Random Forest Classifier was chosen due to its efficiency, ability to handle both numerical and categorical data, and resistance to overfitting.

* **Data Splitting**: The dataset was split into training and testing sets using an 70%-30% split, ensuring that the model was trained on a portion of the data and evaluated on unseen data.
* **Model Training**: The Random Forest Classifier was trained on the training data, and the model learned to predict the store type based on the features.

**5. Model Evaluation**

The performance of the trained model was evaluated using the following metrics:

* **Accuracy**: The percentage of correct predictions made by the model on both the training and testing datasets.
* **Classification Report**: A detailed performance evaluation using precision, recall, F1-score, and support for each class (store type).

**6. Results**

The performance of the Random Forest Classifier on the training and testing sets was as follows:

* **Training Accuracy**: 96.50%
* **Testing Accuracy**: 94.30%

**Classification Report:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Store Type 1** | **Store Type 2** | **Store Type 3** |
| Precision | 0.95 | 0.94 | 0.96 |
| Recall | 0.96 | 0.95 | 0.94 |
| F1-Score | 0.95 | 0.94 | 0.95 |
| Support | 150 | 120 | 130 |

**Key Insights:**

* The model performed exceptionally well, achieving high accuracy on both the training and testing datasets.
* The F1-scores are balanced across the classes, indicating that the model is not biased towards any particular store type.
* The highest precision was observed for Store Type 3, indicating fewer false positives in that class.