Capstone Project: Historical Structures Classification and Tourism Recommendation Engine

Objective: The goal of this project is to harness the power of machine learning and deep learning to classify historical structures and develop a tourism recommendation engine. This initiative aims to preserve historical heritage, support government agencies in maintenance efforts, and enhance tourist experiences through personalized recommendations.

Part 1: Classification of Historical Structures

Business Scenario:
XYZ Pvt. Ltd., a consulting firm, has been engaged to develop an AI model using TensorFlow that can predict the category of a historical structure from images. This model will help monitor these structures' condition and notify agencies about necessary maintenance

Dataset Overview:

- **Training Dataset**: Structures_dataset.zip – Contains images of historical structures categorized into classes.
- **Test Dataset**: dataset_test – Contains images of historical structures for testing.

Project Tasks:

1. **Visualizing Sample Images:**
    - Plot 8–10 sample images from each class using the OpenCV library to understand the dataset better.
2. **Model Selection and Transfer Learning:**
    - Select a suitable CNN architecture for transfer learning.
    - Configure the TensorFlow environment with a chosen backbone and load pre-trained weights.
    - Freeze convolutional layers to retain pre-trained features.
3. **Model Customization:**
    - Modify the top layers of the CNN architecture:
        - Add dense layers with activation functions.
        - Use dropout layers for regularization to prevent overfitting.
    - Tune hyperparameters for optimal performance.
4. **Model Compilation:**
    - Compile the model with appropriate parameters:
        - **Optimizer**: Select an appropriate optimization algorithm (e.g., Adam or SGD).
        - **Loss Function**: Use categorical cross-entropy for multi-class classification.
        - **Metrics**: Monitor accuracy during training.
5. **Training and Evaluation:**
    - Train the model in two phases:
        1. Without data augmentation to establish a baseline.
        2. With data augmentation to improve generalization.
    - Monitor validation accuracy during training using callbacks, such as early stopping.
6. **Visualization:**
    - Plot training and validation accuracy over epochs to assess performance and detect overfitting.

Part 2: Tourism Recommendation Engine

Business Scenario:
In addition to classification, the project requires building a recommendation engine to help tourists explore locations of interest. By analyzing user preferences and ratings, the system will personalize recommendations for enhanced tourist satisfaction.

Dataset Overview:
1. **user.csv**: Contains user demographic data, such as user ID, age, and location.
2. **tourism_with_id.csv**: Details about tourist attractions, including their name, category, and city.
3. **tourism_rating.csv**: Records user ratings for various tourist places.

Project Tasks:
1. **Data Preprocessing:**
    - Inspect datasets for missing values and duplicates.
    - Address anomalies to ensure data integrity.
2. **Exploratory Data Analysis (EDA):**
    - Analyze user demographics:
        - Age distribution of users providing ratings.
        - Origins of users visiting and rating places.
    - Explore locations and categories:
        - Identify various categories of tourist spots.
        - Assess city-based suitability for specific interests (e.g., nature enthusiasts).
3. **Combined Data Creation:**
    - Merge datasets to form a unified view:
        - Identify the most loved spots and categories.
        - Highlight cities with the most popular tourist destinations.
4. **Recommender System Development:**
    - Build a collaborative filtering model using the combined dataset.
    - Personalize recommendations based on user preferences and current location.

Tools and Techniques

1. **Deep Learning**:
    - TensorFlow/Keras for model development.
    - Transfer learning using pre-trained CNN architectures (e.g., ResNet, VGG16).
    - Data augmentation techniques to improve model robustness.
2. **Data Science**:
    - Pandas and NumPy for data preprocessing.
    - Matplotlib and Seaborn for data visualization.
3. **Recommendation Systems**:
    - Collaborative filtering for personalized suggestions.
    - Evaluation of recommendations using metrics like precision and recall.

Expected Outcomes

1. A robust deep learning model capable of accurately classifying historical structures based on their images.
2. Insights into tourism trends, popular locations, and user preferences.
3. A functional recommendation engine that suggests places based on user ratings and interests.

This project combines computer vision and data science to deliver an impactful solution for heritage preservation and tourism enhancement, ensuring that historical and cultural landmarks remain a significant part of our collective history.