AIML Capstone

Historical Structures Classification

Business Scenario

Problem statement 1:

There are hundreds of years-old historical structures that preserve a country's and community's history for future generations and promote tourism opportunities.

To help the travel and tourism industries, it has been decided to use advanced machine learning techniques to monitor the condition of these historical structures and report to government agencies if any of them need maintenance.

Also, understanding customers (tourists) and their expectations is critical for effective marketing. A recommendation engine is an excellent way to supplement existing marketing outreach to prospects.

Business Scenario

This problem objective consists of two parts: Part 1 and Part 2

Part 1:

XYZ Pvt. Ltd., a leading industry consulting firm, has been hired to help the cause by developing an intelligent and automated AI model using TensorFlow that can predict the category of a structure in an image.

Part 2:

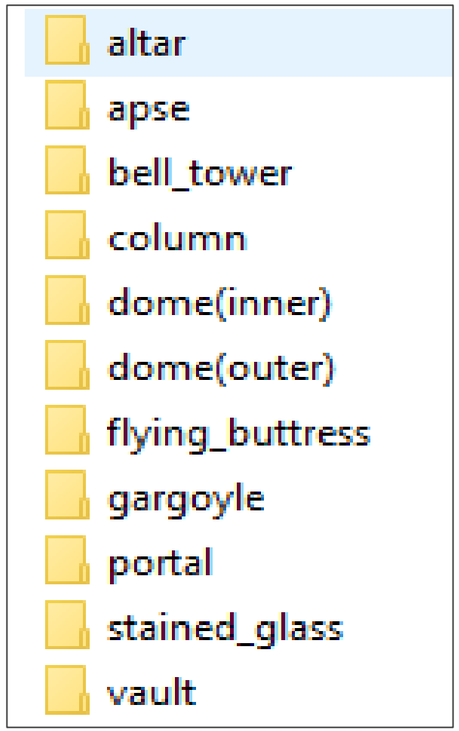
The second objective of this project requires you to perform exploratory data analysis and develop a recommendation engine that will help tourists visit their places of interest.

Part 1

Dataset Snapshot

Name of the dataset: **Structures_dataset.zip**

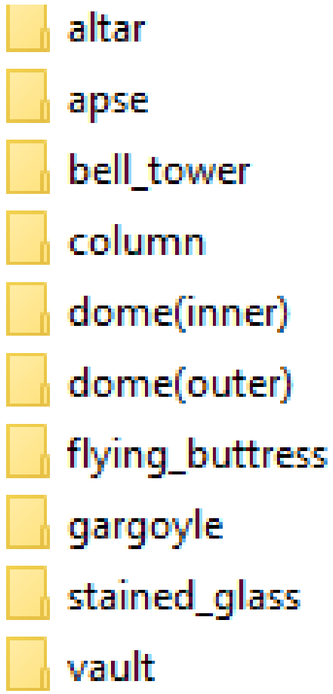
This dataset is the training set and consists of images of historical structure.



Dataset Snapshot

Name of the dataset: **dataset_test**

This dataset is a test set and consists of images of historical structure.



Project Task

Deep learning

1. Plot the sample images (8–10) from each class or category to gain a better understanding of each class

Hint: You can use the OpenCV open-source library for this task.

2. Select an CNN architecture of your choice to train the CV model. Configure the architecture for transfer learning, set up a TensorFlow environment for the selected backbone architecture, and load pre-trained weights

Note: Algorithm or architecture selection is an important step in the training of ML models, so select the one that performs the best on your dataset.

Project Task

Deep learning

3. Deep learning models tend to work well with large amounts (millions of images) of data, but we may not always have enough data to train a deep learning model from scratch. Transfer learning techniques allow us to train and fine-tune large deep learning architectures using limited data.

Hint: For transfer learning, use pre-trained CNN weights and freeze all convolutional layers' weights.

4. As of now, CNN architecture has been configured for our model. Modify the top of this architecture to work with our dataset by:
   - Adding an appropriate number of dense layers with an activation function.
   - Using dropout for regularization.

Note: It is important to understand that these parameters are hyperparameters that must be tuned.

Project Task

Deep learning

5. Compile the model with the right set of parameters like optimizer, loss function, and metric
6. Define your callback class to stop the training once validation accuracy reaches a certain number of your choice
7. Setup the train or test dataset directories and review the number of image samples for the train and test datasets for each class
8. Train the model without augmentation while continuously monitoring the validation accuracy

Project Task

Deep learning

10. Train the model with augmentation and keep monitoring validation accuracy

Note: Choose carefully the number of epochs, steps per epoch, and validation steps based on your computer configuration

11. Visualize training and validation accuracy on the y-axis against each epoch on the x-axis to see if the model overfits after a certain epoch

Part 2

Dataset Snapshot

Name of the dataset: **user.csv**

User_Id	Location	Age
1	Semarang	20
2	Bekasi, Ja	21
3	Cirebon, J	23
4	Bekasi, Ja	21
5	Lampung,	20
6	Jakarta Ut	18
7	Jakarta Se	39
8	Bandung,	40
9	Surabaya,	38
10	Bekasi, Ja	39
11	Yogyakart	20
12	Bogor, Jav	37
13	Depok, Ja	18
14	Jakarta Pu	26
15	Jakarta Tir	34
16	Bekasi, Ja	40
17	Semarang	31
18	Yogyakart	39
19	Cirebon, J	26

Dataset Snapshot

Name of the dataset: **tourisim_with_id.csv**

Place_Id	Place_Name	Description	Category	City	Price	Rating	Time_Minute	Coordinate	Lat	Long		
1	Monumen Na	Monumen Na	Budaya	Jakarta	20000	4.6	15	{'lat': -6.1753	-6.1753924	106.827153		1
2	Kota Tua	Kota tua di Ja	Budaya	Jakarta	0	4.6	90	{'lat': -6.1376	-6.1376448	106.817125		2
3	Dunia Fantas	Dunia Fantas	Taman Hibur	Jakarta	270000	4.6	360	{'lat': -6.1253	-6.1253124	106.833538		3
4	Taman Mini I	Taman Mini I	Taman Hibur	Jakarta	10000	4.5		{'lat': -6.3024	-6.3024459	106.895156		4
5	Atlantis Wate	Atlantis Wate	Taman Hibur	Jakarta	94000	4.5	60	{'lat': -6.1241	-6.12419	106.839134		5
6	Taman Impia	Taman Impia	Taman Hibur	Jakarta	25000	4.5	10	{'lat': -6.1173	-6.1173332	106.857995		6
7	Kebun Binata	Kebun Binata	Cagar Alam	Jakarta	4000	4.5		{'lat': -6.3124	-6.3124593	106.820187		7
8	Ocean Ecopa	Ocean Ecopa	Taman Hibur	Jakarta	180000	4		{'lat': -6.1258	-6.1258017	106.836325		8
9	Pelabuhan M	Pelabuhan M	Bahari	Jakarta	175000	4.4		{'lat': 1.07888	1.07888	103.931398		9
10	Pulau Tidung	Pulau Tidung	Bahari	Jakarta	150000	4.5		{'lat': -5.8032	-5.8032053	106.523791		10
11	Pulau Bidada	Pulau Bidada	Bahari	Jakarta	5000	4.6		{'lat': -6.0358	-6.035833	106.746944		11
12	Pulau Pari	Pulau Pari ad	Bahari	Jakarta	150000	4		{'lat': -5.9074	-5.9074328	106.586399		12
13	Pulau Pramul	Pulau Pramul	Bahari	Jakarta	5000	4.2		{'lat': -5.7459	-5.745962	106.613658		13
14	Pulau Pelang	Pulau Pelang	Bahari	Jakarta	900000	4.8		{'lat': -5.5870	-5.587055	106.5885		14
15	Pasar Seni	Pasar Seni m	Pusat Perbela	Jakarta	0	4.4	90	{'lat': -6.1766	-6.1766868	106.841767		15
16	Jembatan Kot	Jembatan Kot	Budaya	Jakarta	0	4.3		{'lat': -6.1314	-6.1314572	106.810617		16
17	Museum Fata	Museum Fata	Budaya	Jakarta	5000	4.4		{'lat': -6.1364	-6.1364489	106.813066		17
18	Museum Ban	Museum Ban	Budaya	Jakarta	2000	4.7		{'lat': -6.1371	-6.137127	106.813005		18
19	Kidzania	KidZania ada	Taman Hibur	Jakarta	185000	4.6	300	{'lat': -6.2250	-6.2250735	106.809714		19
20	Museum Tam	Museum Tam	Budaya	Jakarta	2000	4.5	90	{'lat': -6.1722	-6.1722241	106.81897		20
21	Museum Wa	Museum Wa	Budaya	Jakarta	5000	4.5	150	{'lat': -6.1349	-6.1349072	106.812445		21
22	Masjid Istiqla	Masjid Istiqla	Tempat Ibada	Jakarta	0	4.7		{'lat': -6.1701	-6.17017	106.83139		22
23	Gereja Kated	Gereja Kated	Tempat Ibada	Jakarta	0	4.8		{'lat': -6.1692	-6.169225	106.833063		23

Dataset Snapshot

Name of the dataset: **tourism_rating.csv**

User_Id	Place_Id	Place_Ratings
1	179	3
1	344	2
1	5	5
1	373	3
1	101	4
1	312	2
1	258	5
1	20	4
1	154	2
1	393	5
1	103	3
1	208	5
1	89	3
1	405	5
1	41	5
1	336	4
1	67	4
1	292	3
1	222	3

Dataset Description

user.csv: It includes user demographic data to help with recommendations.

Variables	Description
User_id	Unique identification for the user
location	Location the user has visited
age	Age of the visitor

Dataset Description

Toursim_with_id.csv: It provides details on the tourist attractions in Indonesia's five largest cities.

Variables	Description
Place_id	Unique identification for the place
Place_name	Name of the place
Description	Description of the place
Category	Category of the place
City	City for the tourist destination
Price	Price of touring the place
Rating	Rating of the destination
Time_minute	Time in minutes taken to tour the place
Coordinate	Coordinates of the place
Lat	Latitude of the location
Long	Longitude of the location

Dataset Description

Tourism_rating.csv: It has three columns: the user, the location, and the rating, which is used to build a recommendation engine based on the rating.

Variables	Description
user_id	The user id of the visitor
place_id	Unique identification number of the place
place_rating	Rating of the place

Project Task

Data science

1. Import all the datasets and perform preliminary inspections, such as:
    I. Check for missing values and duplicates
    II. Remove any anomalies found in the data
2. To understand the tourism highlights better, we should explore the data in depth.
    I. Explore the user group that provides the tourism ratings by:
        - Analyzing the age distribution of users visiting the places and rating them
        - Identifying the places where most of these users (tourists) are coming from

Project Task

Data science

3. Next, let's explore the locations and categories of tourist spots.
    I. What are the different categories of tourist spots?
    II. What kind of tourism each location is most famous or suitable for?
    III. Which city would be best for a nature enthusiast to visit?
4. To better understand tourism, we need to create a combined data with places and their user ratings.

Project Task

Data science

I. Use this data to figure out the spots that are most loved by the tourists. Also, which city has the most loved tourist spots?

II. Indonesia provides a wide range of tourist spots ranging from historical and cultural beauties to advanced amusement parks. Among these, which category of places are users liking the most?

5. Build a recommender model for the system

I. Use the above data to develop a collaborative filtering model for recommendation and use that to recommend other places to visit using the current tourist location(place name)

Thank You