Curated AIML Capstone Projects

Content:

1. **Agri Analytics**
2. **Optical Character Recognition**
3. **Financial Analytics**
4. **Healthcare Analytics**
5. **Image Analytics**
6. **Retail & Supply Chain Analytics**
7. **Natural Language Processing**
8. **Reinforcement Learning based projects**
9. **AI Products**
10. **Legal Analytics**

Important Information: In the following sections, we list some of the problems students can take as capstone project. Observe that in any data science project, 50% to 75% of the time, data scientists will be spending time on data collection, curation and exploratory data analysis on the collected data. Along with problem definitions, for some problems, we have also provided data sources links. Great Learning does not provide data to execute capstone project. Students need to collect data on their own either following the links provided or from some other sources. Students can take their Mentor's suggestions in data collection. Observe that Mentor will provide guidance in project execution, validating the results (excluding the code debugging) & validating the project related documents (written by students). Students need to have at least one meeting (30 minutes' duration) for a week with Mentor.
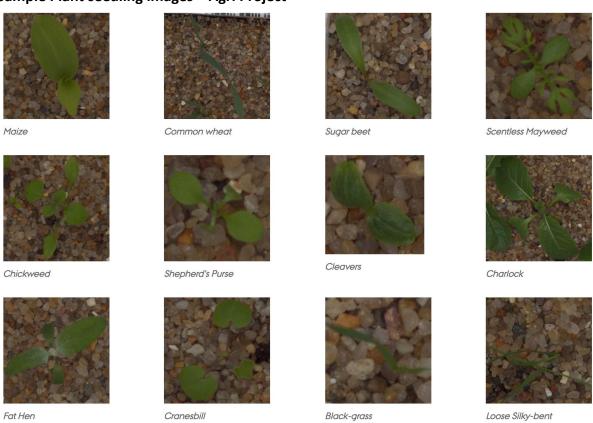
1. Agri Analytics

Crop Price prediction
Portfolio of crops
Wastage minimisation
Supply chain
Weed identification
Yield prediction
Agri chat Bot
Precision agriculture

Data Sources : https://catalog.data.gov/dataset?groups=agriculture8571
https://lionbridge.ai/datasets/14-free-agriculture-datasets-for-machine-learning/

Sample Plant seedling images – Agri Project

Maize

Common wheat

Sugar beet

Scentless Mayweed

Chickweed

Shepherd's Purse

Cleavers

Charlock

Fat Hen

Cranesbill

Black-grass

Loose Silky-bent

2. Optical character recognition (OCR)

Standard form (Hand Written) data extraction (Pan application, Insurance form)
Standard form (Printed) data extraction
Non-standard from data extraction
Non-standard form hand written

Data Sources : https://lionbridge.ai/datasets/15-best-ocr-handwriting-datasets/

Sample of OCR

3. Financial Analytics

Direction prediction
AI + fundamental based stock prediction
Peer to Peer lending and credit default prediction
   a. Data set: Lending club data set
Prediction and classification of non-performing assets
Financial markets
   a. Predict market direction using social media data

   b. Consider 100 top market value stocks, apply 101 alphas from worldQuant and build robust trading strategies by using ensemble techniques or reinforcement learning. Construct diversified portfolio. References for this are as follows:

> ➢ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2701346
> ➢ https://www.quantopian.com/posts/the-101-alphas-project

Chat-Bot for Insurance Industry
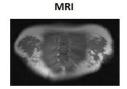Chat-Bot for Peer to Peer lending

4. Healthcare Analytics

Part	Modalities	Dimensions
Chest	X-ray	2D
Liver	CT - Scan	3D
Brain	MRI	
Breast		
Bones		
Abdominal		

Sample Images

Chest Images   X-ray   CT - Scan   MRI

a. Pathology detection using Chest X-rays
> ➢ NIH dataset = 112120 chest x-rays with 14 pathology labels from
> https://nihcc.app.box.com/v/ChestXray-NIHCC
b. Pathology location detection in Chest X-rays (bounding boxes)
> ➢ https://www.kaggle.com/c/rsna-pneumonia-detection-challenge
c. Automatic reports generation using Chest X-rays and historical reports
> ➢ https://openi.nlm.nih.gov/faq.php
d. Identify metastatic tissue in histopathologic scans of lymph node sections
> ➢ https://www.kaggle.com/c/histopathologic-cancer-detection
e. Classify subcellular protein patterns in human cells
> ➢ https://www.kaggle.com/c/human-protein-atlas-image-classification
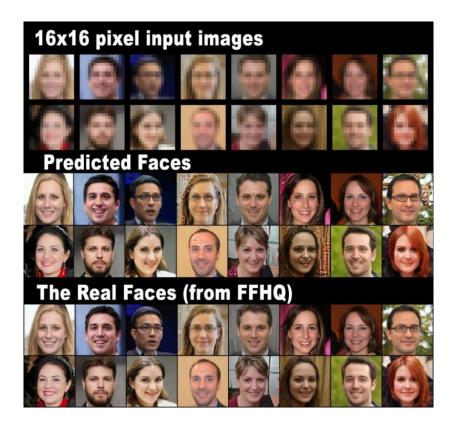
5. Image Analytics

Types of Projects

Photo
Video
Photo + Text
Video + Text

a. Image Classification of Furniture & Home Good

➢ https://www.kaggle.com/c/imaterialist-challenge-furniture-2018

b. Given an image, can you find all of the same landmarks in a dataset?

➢ https://www.kaggle.com/c/landmark-retrieval-challenge

c. Identify and classify toxic online comments

➢ https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data

d. Identify and label everyday objects in images

➢ https://www.kaggle.com/c/imagenet-object-detection-challenge

e. Determine the species of a seedling from an image

   ➢ https://www.kaggle.com/c/plant-seedlings-classification/data

f. Develop algorithms to predict the adoptability of pets - specifically, how quickly is a pet adopted

   ➢ https://www.kaggle.com/c/petfinder-adoption-prediction

g. Can you predict upcoming laboratory earthquakes?

   ➢ https://www.kaggle.com/c/LANL-Earthquake-Prediction/data

h. Power line fault detection

   ➢ https://www.kaggle.com/c/vsb-power-line-fault-detection

i. Identify objects in images

   ➢ https://www.kaggle.com/c/imagenet-object-localization-challenge

j. Determine whether an image contains a columnar cactus

   ➢ https://www.kaggle.com/c/aerial-cactus-identification

k. Find ships on satellite images as quickly as possible

   ➢ https://www.kaggle.com/c/airbus-ship-detection

l. Segment salt deposits beneath the Earth's surface

   ➢ https://www.kaggle.com/c/tgs-salt-identification-challenge

m. Image classification of fashion products

   ➢ https://www.kaggle.com/c/imaterialist-challenge-fashion-2018

n. Given an image, can you find all of the same landmarks in a dataset?

   ➢ https://www.kaggle.com/c/landmark-retrieval-challenge

o. Label famous (and not-so-famous) landmarks in images

   ➢ https://www.kaggle.com/c/landmark-recognition-challenge

p. Determine the breed of a dog in an image

   ➢ https://www.kaggle.com/c/dog-breed-identification

6. Retail & Supply chain Analytics

Customer segmentation & Churn prediction
Internet of things (IOT)
Chat BOT – Supply chain
Supply chain optimization

Retail Projects:

c. Retail prediction - Mercari Price Suggestion Challenge, Objective: build an algorithm that automatically suggests the right product prices
   ➢ https://www.kaggle.com/c/mercari-price-suggestion-challenge

d. Predict sales prices and practice feature engineering, RFs, and gradient boosting
   ➢ https://www.kaggle.com/c/house-prices-advanced-regression-techniques

e. Forecast the total amount of products sold in every shop
   ➢ https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data

f. Predict future sales
   ➢ https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data

g. Identify which customers will make a specific transaction in the future, irrespective of the amount of money transacted
   ➢ https://www.kaggle.com/c/santander-customer-transaction-prediction

7. Natural Language Processing

Language Translation
Chat BOT
Summarization

a. Identify and classify toxic online comments

   https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data

8. RL Based projects

Robotics
Supply chain

9. AI products

Driver distraction
Scanning by mobile & OCR extraction

10. Legal

Case outcome prediction