

30

Deep Learning Projects

With

Datasets Details

Covid19

- Bing Coronavirus
 - Classify Bing Queries as either specific (e.g. about a specific location) or generic. You might have to figure out a more exact definition of specific or generic though
 - Dataset: [BingCoronavirusQuerySet](#)
- Covid Clinical Data
 - Rank and sort high risk patients using clinical data. Pick an interpretable approach if you can.
 - Dataset: [CovidClinicalData](#)

If you haven't already, checkout [Kaggle's Covid19 Section](#) as well. It has datasets and ideas both.

Text

- Autonomous Tagging of StackOverflow Questions
 - Make a multi-label classification system that automatically assigns tags for questions posted on a forum such as StackOverflow or Quora.
 - Dataset: [StackLite](#) or [10% sample](#)
- Keyword/Concept identification
 - Identify keywords from millions of questions
 - Dataset: [StackOverflow question samples by Facebook](#)
- Topic identification
 - Multi-label classification of printed media articles to topics
 - Dataset: [Greek Media monitoring multi-label classification](#)

Natural Language Understanding

- Sentence to Sentence semantic similarity
 - Can you identify question pairs that have the same intent or meaning?
 - Dataset: [Quora question pairs](#) with similar questions marked
- Fight online abuse
 - Can you confidently and accurately tell whether a particular comment is abusive?
 - Dataset: [Toxic comments on Kaggle](#)
- Open Domain question answering
 - Can you build a bot which answers questions according to the student's age or her curriculum?
 - [Facebook's FAIR](#) is built in a similar way for Wikipedia.
 - Dataset: [NCERT books](#) for K-12/school students in India, [NarrativeQA by Google DeepMind](#) and [SQuAD by Stanford](#)
- Automatic text summarization
 - Can you create a summary with the major points of the original document?
 - Abstractive (write your own summary) and Extractive (select pieces of text from original) are two popular approaches
 - Dataset: [CNN and DailyMail News Pieces](#) by Google DeepMind
- Copy-cat Bot
 - Generate plausible new text which looks like some other text
 - Obama Speeches? For instance, you can create a bot which writes some [new speeches in Obama's style](#)
 - Trump Bot? Or a Twitter bot which mimics [@realDonaldTrump](#)
 - Narendra Modi bot saying "*doston*"? Start by scrapping off his *Hindi* speeches from his [personal website](#)
 - Example Dataset: [English Transcript of Modi speeches](#)

Check [mlm/blog](#) for some hints.

- Sentiment Analysis
 - Do Twitter Sentiment Analysis on tweets sorted by geography and timestamp.
 - Dataset: [Tweets sentiment tagged by humans](#)

Forecasting

- Univariate Time Series Forecasting
 - How much will it rain this year?

- Dataset: [45 years of rainfall data](#)
- Multi-variate Time Series Forecasting
 - How polluted will your town's air be? Pollution Level Forecasting
 - Dataset: [Air Quality dataset](#)
- Demand/load forecasting
 - Find a short term forecast on electricity consumption of a single home
 - Dataset: [Electricity consumption of a household](#)
- Predict Blood Donation
 - We're interested in predicting if a blood donor will donate within a given time window.
 - More on the problem statement at [Driven Data](#).
 - Dataset: [UCI ML Datasets Repo](#)

Recommendation systems

- Movie Recommender
 - Can you predict the rating a user will give on a movie?
 - Do this using the movies that user has rated in the past, as well as the ratings similar users have given similar movies.
 - Dataset: [Netflix Prize](#) and [MovieLens Datasets](#)
- Search + Recommendation System
 - Predict which Xbox game a visitor will be most interested in based on their search query
 - Dataset: [BestBuy](#)
- Can you predict Influencers in the Social Network?
 - How can you predict social influencers?
 - Dataset: [PeerIndex](#)

Vision

- Image classification
 - Object recognition or image classification task is how Deep Learning shot up to it's present-day resurgence
 - Datasets:
 - [CIFAR-10](#)
 - [ImageNet](#)
 - [MS COCO](#) is the modern replacement to the ImageNet challenge

- [MNIST Handwritten Digit Classification Challenge](#) is the classic entry point
 - [Character recognition \(digits\)](#) is the good old Optical Character Recognition problem
 - Bird Species Identification from an Image using the [Caltech-UCSD Birds dataset](#) dataset
- Diagnosing and Segmenting Brain Tumors and Phenotypes using MRI Scans
 - Dataset: MICCAI Machine Learning Challenge aka [MLC 2014](#)
- Identify endangered right whales in aerial photographs
 - Dataset: [MOAA Right Whale](#)
- Can computer vision spot distracted drivers?
 - Dataset: [State Farm Distracted Driver Detection](#) on Kaggle
- Bone X-Ray competition
 - Can you identify if a hand is broken from a X-ray radiographs automatically with better than human performance?
 - Stanford's Bone XRay Deep Learning Competition with [MURA Dataset](#)
- Image Captioning
 - Can you caption/explain the photo a way human would?
 - Dataset: [MS COCO](#)
- Image Segmentation/Object Detection
 - Can you extract an object of interest from an image?
 - Dataset: [MS COCO](#), [Carvana Image Masking Challenge](#) on Kaggle
- Large-Scale Video Understanding
 - Can you produce the best video tag predictions?
 - Dataset: [YouTube 8M](#)
- Video Summarization
 - Can you select the semantically relevant/important parts from the video?
 - Example: [Fast-Forward Video Based on Semantic Extraction](#)
 - Dataset: Unaware of any standard dataset or agreed upon metrics? I think [YouTube 8M](#) might be good starting point.
- Style Transfer
 - Can you recompose images in the style of other images?
 - Dataset: [fzliu on GitHub](#) shared target and source images with results

- Chest XRay
 - Can you detect if someone is sick from their chest XRay? Or guess their radiology report?
 - Dataset: [MIMIC-CXR at Physionet](#)
- Clinical Diagnostics: Image Identification, classification & segmentation
 - Can you help build an open source software for lung cancer detection to help radiologists?
 - Link: [Concept to clinic](#) challenge on DrivenData
- Satellite Imagery Processing for Socioeconomic Analysis
 - Can you estimate the standard of living or energy consumption of a place from night time satellite imagery?
 - Reference for Project details: [Stanford Poverty Estimation Project](#)
- Satellite Imagery Processing for Automated Tagging
 - Can you automatically tag satellite images with human features such as buildings, roads, waterways and so on?
 - Help free the manual effort in tagging satellite imagery: [Kaggle Dataset by DSTL, UK](#)

Music

- Music/Audio Recommendation Systems
 - Can you tell if two songs are similar using their sound or lyrics?
 - Dataset: [Million Songs Dataset](#) and it's 1% sample.
 - Example: [Anusha et al](#)
- Music Genre recognition using neural networks
 - Can you identify the musical genre using their spectrograms or other sound information?
 - Datasets: [FMA](#) or [GTZAN on Keras](#)
 - Get started with [Librosa](#) for feature extraction

Data Science ML Full Stack Roadmap

<https://github.com/hemansnation/Data-Science-ML-Full-Stack-2022>

**Join the Data Science & ML Full Stack WhatsApp Group Community here:
If the group is full, please join another one.**

<https://chat.whatsapp.com/B7Mdp6QTMJ0KZYGWrzIT3Y>
<https://chat.whatsapp.com/HWDSJU4KXrXJlcn5Npp3Gm>
<https://chat.whatsapp.com/DmATV5uaVY7IKrTMHDiHnr>
<https://chat.whatsapp.com/Blz2n8QYSgdKWfQbJZxHtJ>

Join Telegram for Data Science ML AI Resources:

<https://t.me/+sREuRiFssMo4YWJl>

Join Community on LinkedIn:

<https://www.linkedin.com/groups/12540639/>

Connect with me on these platforms:

LinkedIn: <https://www.linkedin.com/in/hemansnation/>

Twitter: <https://twitter.com/hemansnation>

GitHub: <https://github.com/hemansnation>

Instagram: <https://www.instagram.com/masterdexter.ai/>

Are you a professional?

DM for One-on-One sessions for Python, Data Science, Machine Learning, and Data Engineering.

Here: <https://bit.ly/3U6zQvQ>