Machine Learning

Public Datasets

Data Processing X Machine Learning

Datasets for your next project in Data Science, Machine Learning, Deep Learning domain and sector wise.

Himanshu Ramchandani
https://www.linkedin.com/in/hemansnation/

Contents

Museums

NaturalLanguage

Neuroscience

Physics

ProstateCancer

Psychology+Cognition

PublicDomains

SearchEngines

SocialNetworks

SocialSciences

Software

Sports

TimeSeries

Transportation

eSports

Complementary Collections

Agriculture

- ✅ The global dataset of historical yields for major crops 1981–2016 - The Global Dataset of
- ✅ Hyperspectral benchmark dataset on soil moisture - This dataset was measured in a five-day
- ✅ Lemons quality control dataset - Lemon dataset has been prepared to investigate the
- ✅ Optimized Soil Adjusted Vegetation Index - The IDB is a tool for working with remote sensing
- ✅ U.S. Department of Agriculture's Nutrient Database
- ✅ U.S. Department of Agriculture's PLANTS Database - The Complete PLANTS Checklist is nearly 7

Architecture

- ✅ Swiss Apartment Models - This dataset contains detailed data on 42,207 apartments (242,257

Biology

- ✅ 1000 Genomes - The 1000 Genomes Project ran between 2008 and 2015, creating the largest
- ✅ American Gut (Microbiome Project) - The American Gut project is the largest crowdsourced
- ✅ BCNB - There are WSIs of 1058 patients, part of tumor regions are annotated in WSIs. Except
- ✅ Broad Bioimage Benchmark Collection (BBBC) - The Broad Bioimage Benchmark Collection (BBBC)
- ✅ Broad Cancer Cell Line Encyclopedia (CCLE)
- ✅ Cell Image Library - This library is a public and easily accessible resource database of
- ✅ Complete Genomics Public Data - A diverse data set of whole human genomes are freely

- ✓ CytoImageNet - A large-scale dataset of microscopy images. Contains 890,737 total grayscale
- ✓ EBI ArrayExpress - ArrayExpress Archive of Functional Genomics Data stores data from high-
- ✓ EBI Protein Data Bank in Europe - The Electron Microscopy Data Bank (EMDB) is a public
- ✓ ENCODE project - The Encyclopedia of DNA Elements (ENCODE) Consortium is an ongoing
- ✓ Electron Microscopy Pilot Image Archive (EMPIAR) - EMPIAR, the Electron Microscopy Public
- ✓ Ensembl Genomes
- ✓ Gene Expression Omnibus (GEO) - GEO is a public functional genomics data repository
- ✓ Gene Ontology (GO) - GO annotation files
- ✓ Global Biotic Interactions (GloBI)
- ✓ Harvard Medical School (HMS) LINCS Project - The Harvard Medical School (HMS) LINCS Center is
- ✓ Human Genome Diversity Project - A group of scientists at Stanford University have
- ✓ Human Microbiome Project (HMP) - The HMP sequenced over 2000 reference genomes isolated from
- ✓ ICOS PSP Benchmark - The ICOS PSP benchmarks repository contains an adjustable real-world
- ✓ International HapMap Project
- ? Journal of Cell Biology DataViewer
- ✓ KEGG - KEGG is a database resource for understanding high-level functions and utilities of
- ✓ NCBI Proteins
- ✓ NCBI Taxonomy - The NCBI Taxonomy database is a curated set of names and classifications for
- ✓ NCI Genomic Data Commons - The GDC Data Portal is a robust data-driven platform that allows
- ✓ NIH Microarray data
- ✓ OpenSNP genotypes data - openSNP allows customers of direct-to-customer genetic tests to

- ✅ Palmer Penguins - The goal of palmerpenguins is to provide a great dataset for data
- ✅ Pathguid - Protein-Protein Interactions Catalog
- ✅ Protein Data Bank - This resource is powered by the Protein Data Bank archive-information
- ✅ Psychiatric Genomics Consortium - The purpose of the Psychiatric Genomics Consortium (PGC) is
- ✅ PubChem Project - PubChem is the world's largest collection of freely accessible chemical
- ✅ PubGene (now Coremine Medical) - COREMINE™ is a family of tools developed by the Norwegian
- ✅ Sanger Catalogue of Somatic Mutations in Cancer (COSMIC) - COSMIC, the Catalogue Of Somatic
- ✅ Sanger Genomics of Drug Sensitivity in Cancer Project (GDSC)
- ✅ Sequence Read Archive(SRA) - The Sequence Read Archive (SRA) stores raw sequence data from
- ✅ Serratus - Analysis of 7.1 million RNA/DNA sequencing datasets to discover the total
- ✅ Stanford Microarray Data (Retired NOW)
- ✅ Stowers Institute Original Data Repository
- ✅ Systems Science of Biological Dynamics (SSBD) Database - Systems Science of Biological
- ✅ The Cancer Genome Atlas (TCGA), available via Broad GDAC
- ✅ The Catalogue of Life - The Catalogue of Life is a quality-assured checklist of more than 1.8
- ✅ The Personal Genome Project - The Personal Genome Project, initiated in 2005, is a vision and
- ✅ UCSC Public Data
- ✅ UniGene
- ✅ Universal Protein Resource (UnitProt) - The Universal Protein Resource (UniProt) is a
- ✅ Rfam - The Rfam database is a collection of RNA families, each represented by multiple

Chemistry

- ✅ Ionic Liquids Database - ILThermo

Climate+Weather

- ❓ Actuaries Climate Index
- ❓ Australian Weather
- ✅ Aviation Weather Center - Consistent, timely and accurate weather information for the world
- ✅ Brazilian Weather - Historical data (In Portuguese) - Data related to climate and weather
- ✅ Canadian Meteorological Centre
- ❓ Climate Data from UEA (updated monthly)
- ✅ Dutch Weather - The KNMI Data Center (KDC) portal provides access to KNMI data on weather,
- ✅ European Climate Assessment & Dataset
- ✅ German Climate Data Center
- ✅ Global Climate Data Since 1929
- ✅ Charting The Global Climate Change News Narrative 2009-2020 - These four datasets represent
- ✅ NASA Global Imagery Browse Services
- ✅ NOAA Bering Sea Climate
- ✅ NOAA Climate Datasets
- ❓ NOAA Realtime Weather Models
- ✅ NOAA SURFRAD Meteorology and Radiation Datasets
- ✅ The World Bank Open Data Resources for Climate Change
- ❓ UEA Climatic Research Unit
- ✅ WU Historical Weather Worldwide
- ✅ Wahington Post Climate Change - To analyze warming temperatures in the United States, The
- ✅ WorldClim - Global Climate Data

ComplexNetworks

- ✓ AMiner Citation Network Dataset
- ✓ CrossRef DOI URLs
- ✓ DBLP Citation dataset
- ✓ DIMACS Road Networks Collection
- ✓ NBER Patent Citations
- ✓ NIST complex networks data collection
- ? Network Repository with Interactive Exploratory Analysis Tools
- ✓ Protein-protein interaction network
- ✓ PyPI and Maven Dependency Network
- ✓ Scopus Citation Database
- ✓ Small Network Data
- ✓ Stanford GraphBase
- ✓ Stanford Large Network Dataset Collection
- ? Stanford Longitudinal Network Data Sources
- ? The Koblenz Network Collection
- ✓ The Laboratory for Web Algorithmics (UNIMI)
- ? UCI Network Data Repository
- ✓ UFL sparse matrix collection
- ? WSU Graph Database
- ✓ Community Resource for Archiving Wireless Data At Dartmouth - Contains datasets of pcap files

ComputerNetworks

- ✓ 3.5B Web Pages from CommonCrawl 2012
- ✓ 53.5B Web clicks of 100K users in Indiana Univ.
- ✓ CAIDA Internet Datasets
- ? CRAWDAD Wireless datasets from Dartmouth Univ.
- ✓ ClueWeb09 - 1B web pages
- ✓ ClueWeb12 - 733M web pages
- ✓ CommonCrawl Web Data over 7 years

- ✓ Shopper Intent Prediction from Clickstream E-Commerce Data with Minimal Browsing Information
- ✓ Criteo click-through data
- ✓ Internet-Wide Scan Data Repository
- ✓ MIRAGE-2019 - MIRAGE-2019 is a human-generated dataset for mobile traffic analysis with
- ✓ OONI: Open Observatory of Network Interference - Internet censorship data
- ✓ Open Mobile Data by MobiPerf
- ✓ The Peer-to-Peer Trace Archive - Real-world measurements play a key role in studying the
- ✓ Rapid7 Sonar Internet Scans
- ✓ UCSD Network Telescope, IPv4 /8 net

CyberSecurity

- ✓ CCCS-CIC-AndMal-2020 - The dataset includes 200K benign and 200K malware samples totalling to
- ✓ Traffic and Log Data Captured During a Cyber Defense Exercise - This dataset was acquired

DataChallenges

- ✓ AIcrowd Competitions
- ✓ Bruteforce Database
- ✓ Challenges in Machine Learning
- ? CrowdANALYTIX dataX
- ? D4D Challenge of Orange
- ✓ DrivenData Competitions for Social Good
- ✓ ICWSM Data Challenge (since 2009)
- ✓ KDD Cup by Tencent 2012
- ✓ Kaggle Competition Data
- ✓ Localytics Data Visualization Challenge
- ✓ Netflix Prize
- ✓ Space Apps Challenge

- ? Telecom Italia Big Data Challenge
- ? TravisTorrent Dataset - MSR'2017 Mining Challenge
- ✓ TunedIT - Data mining & machine learning data sets, algorithms, challenges
- ✓ Yelp Dataset Challenge - The Yelp dataset is a subset of our businesses, reviews, and user

EarthScience

- ✓ 38-Cloud (Cloud Detection) - Contains 38 Landsat 8 scene images and their manually extracted
- ✓ AQUASTAT - Global water resources and uses
- ✓ BODC - marine data of ~22K vars
- ✓ EOSDIS - NASA's earth observing system data
- ✓ Earth Models
- ✓ Global Wind Atlas - The Global Wind Atlas is a free, web-based application developed to help
- ✓ Integrated Marine Observing System (IMOS) - roughly 30TB of ocean measurements
- ? Marinexplore - Open Oceanographic Data
- ? Alabama Real-Time Coastal Observing System
- ✓ National Estuarine Research Reserves System-Wide Monitoring Program - long-term estuarine
- ✓ Oil and Gas Authority Open Data - The dataset covers 12,500 offshore wellbores, 5,000 seismic
- ✓ Smithsonian Institution Global Volcano and Eruption Database
- ✓ USGS Earthquake Archives
- ✓ Wellhead Protection Area (protection zone) prediction using breakthrough curves - This

Economics

- ✓ Asian Productivity Organization (APO) - The AEPM provides a graphic dashboard view of
- ✓ ASEAN Stats - The ASEANstatsDataPortal was first launched in June 2018. The Portal is
- ✓ American Economic Association (AEA)
- ✓ Asian KLEMS - Asia KLEMS is an Asian regional research consortium to promote building
- ✓ Harvard Atlas of Economic Complexity - A database for people to explore global trade flows
- ✓ BIS Financial Database - The files contain the same data as in the BIS Statistics Explorer
- ✓ Barro-Lee Education Attainment - Barro-Lee Educational Attainment Data from 1950 to 2010.
- ✓ CEPII Database - A database of the world economy, through its country and region profiles, in
- ✓ EUKLEMS - EU KLEMS is an industry level, growth and productivity research project. EU KLEMS
- ✓ Economic Freedom of the World Data
- ✓ Historical National Accounts - The datahub on Comparative Historical National Accounts
- ✓ Historical MacroEconomic Statistics
- ✓ INFORUM - Interindustry Forecasting at the University of Maryland
- ✓ DBnomics – the world's economic database - Aggregates hundreds of millions of time series
- ✓ International Trade Statistics
- ✓ Internet Product Code Database
- ✓ Joint External Debt Data Hub
- ? Jon Haveman International Trade Data Links
- ✓ Latin America KLEMS - LAKLEMS is a technical cooperation project financed by the Inter-
- ✓ Long-Term Productivity Database - The Long-Term Productivity database was created as a
- ✓ Maddison Project Database - The Maddison Project Database provides information on comparative

- ✅ National Transfer Accounts - The goal of the National Transfer Accounts (NTA) project is to
- ✅ OpenCorporates Database of Companies in the World
- ✅ Our World in Data
- ✅ Penn World Table - PWT version 10.0 is a database with information on relative levels of
- ❓ SciencesPo World Trade Gravity Datasets
- ✅ The Atlas of Economic Complexity
- ✅ The Center for International Data
- ✅ The Observatory of Economic Complexity
- ❓ UN Commodity Trade Statistics
- ✅ UN Human Development Reports
- ✅ World Input-Output Database - World Input-Output Tables and underlying data, covering 43
- ✅ World KLEMS - Analytical KLEMS-type data sets for a broad set of countries around the world.

Education

- ✅ College Scorecard Data
- ✅ New York State Education Department Data - The New York State Education Department (NYSED) is
- ❓ Program for International Student Assessement (PISA) - Contains 15-year-old students'
- ✅ Student Data from Free Code Camp

Energy

- ✅ AMPds - The Almanac of Minutely Power dataset
- ✅ BLUEd - Building-Level fUlly labeled Electricity Disaggregation dataset
- ✅ COMBED
- ✅ DBFC - Direct Borohydride Fuel Cell (DBFC) Dataset
- ✅ DEL - Domestic Electrical Load study datsets for South Africa (1994 - 2014)

- ✅ ECO - The ECO data set is a comprehensive data set for non-intrusive load monitoring and
- ✅ EIA
- ✅ Global Power Plant Database - The Global Power Plant Database is a comprehensive, open source
- ✅ HES - Household Electricity Study, UK
- ✅ HFED
- ✅ MORED: a Moroccan Buildings’ Electricity Consumption Dataset - Since spring of 2019, a data
- ✅ Marktstammdatenregister - The German Marktstammdatenregister (MaStR) is a database of all
- ✅ PEM1 - Proton Exchange Membrane (PEM) Fuel Cell Dataset
- ❓ PLAID - The Plug Load Appliance Identification Dataset
- ✅ The Public Utility Data Liberation Project (PUDL) - PUDL makes US energy data easier to
- ❓ REDD
- ✅ SYND - A synthetic energy dataset for non-intrusive load monitoring - With SynD, we present a
- ❓ Smart Meter Data Portal - The Smart Meter Data Portal is part of the National Science
- ✅ Tracebase
- ✅ Ukraine Energy Centre Datasets
- ✅ UK-DALE - UK Domestic Appliance-Level Electricity
- ✅ WHITED
- ✅ iAWE

Entertainment

- ✅ Top Streamers on Twitch - This contains data of Top 1000 Streamers from past year.

Finance

- ✅ BIS Statistics - BIS statistics, compiled in cooperation with central banks and other
- ✅ Blockmodo Coin Registry - A registry of JSON formatted information files that is primarily
- ❓ CBOE Futures Exchange
- ✅ Complete FAANG Stock data - This data set contains all the stock data of FAANG companies from
- ✅ Google Finance
- ✅ Google Trends
- ✅ NASDAQ
- ✅ NYSE Market Data
- ❓ OANDA
- ❓ OSU Financial data
- ✅ Quandl
- ✅ SEC EDGAR - EDGAR, the Electronic Data Gathering, Analysis, and Retrieval system, is the
- ✅ St Louis Federal
- ✅ Yahoo Finance

GIS

- ✅ Awesome 3D Semantic City Models - Collection of open 3D semantic city and region models.
- ✅ ArcGIS Open Data portal
- ✅ Cambridge, MA, US, GIS data on GitHub
- ✅ Database of all continents, countries, States/Subdivisions/Provinces and Cities - Database
- ❓ Factual Global Location Data
- ✅ IEEE Geoscience and Remote Sensing Society DASE Website
- ✅ Geo Maps - High Quality GeoJSON maps programmatically generated
- ❓ Geo Spatial Data from ASU
- ✅ Geo Wiki Project - Citizen-driven Environmental Monitoring
- ✅ GeoFabrik - OSM data extracted to a variety of formats and areas
- ✅ GeoNames Worldwide

- ✓ Global Administrative Areas Database (GADM) - Geospatial data organized by country. Includes
- ✓ Homeland Infrastructure Foundation-Level Data
- ✓ Landsat 8 on AWS
- ✓ List of all countries in all languages
- ✓ National Weather Service GIS Data Portal
- ? Natural Earth - vectors and rasters of the world
- ✓ OpenAddresses
- ✓ OpenStreetMap (OSM)
- ✓ Pleiades - Gazetteer and graph of ancient places
- ✓ Reverse Geocoder using OSM data
- ✓ Robin Wilson - Free GIS Datasets
- ✓ Shadow Accrual Maps - The repository contains the accumulated shadow information for New York
- ✓ TIGER/Line - U.S. boundaries and roads
- ✓ TZ Timezones shapefile
- ✓ TwoFishes - Foursquare's coarse geocoder
- ✓ UN Environmental Data
- ✓ World boundaries from the U.S. Department of State
- ✓ World countries in multiple formats

Government

- ✓ Alberta, Province of Canada
- ? Antwerp, Belgium
- ✓ Argentina (non official)
- ✓ Datos Argentina - Portal de datos abiertos de la República Argentina. Encontrá datos públicos
- ✓ Austin, TX, US
- ✓ Australia (abs.gov.au)
- ✓ Australia (data.gov.au)
- ✓ Austria (data.gv.at)
- ✓ Baton Rouge, LA, US

- ? Beersheba, Israel - Open Data Portal (Smart7 OpenData)
- ✓ Belgium
- ✓ City of Berkeley Open Data
- ? Brazil
- ✓ Buenos Aires, Argentina
- ✓ Calgary, AB, Canada
- ✓ Cambridge, MA, US
- ✓ Canada
- ✓ Chicago
- ? Chile
- ? China
- ✓ Dallas Open Data
- ✓ DataBC - data from the Province of British Columbia
- ✓ Debt to the Penny - The Debt to the Penny dataset provides information about the total
- ✓ Denver Open Data
- ✓ Durham, NC Open Data
- ✓ Edmonton, AB, Canada
- ✓ England LGInform
- ✓ EuroStat
- ✓ EveryPolitician - Ongoing project collating and sharing data on every politician.
- ✓ Federal Committee on Statistical Methodology (FCSM) (formerly FedStats)
- ✓ Finland
- ? France
- ✓ Fredericton, NB, Canada
- ✓ Gatineau, QC, Canada
- ✓ Germany
- ✓ Ghent, Belgium
- ✓ Glasgow, Scotland, UK
- ✓ Greece
- ✓ Guardian world governments
- ? Halifax, NS, Canada

- ✅ Helsinki Region, Finland
- ✅ Hong Kong, China
- ✅ Houston, TX, US
- ✅ Indian Government Data
- ✅ Indonesian Data Portal
- ✅ Iowa - Welcome to the State of Iowa's data portal. Please explore data about Iowa and your
- ✅ Ireland's Open Data Portal
- ✅ Israel's Open Data Portal
- ❓ Istanbul Municipality Open Data Portal
- ✅ Italy - Il Portale dati.gov.it è il catalogo nazionale dei metadati relativi ai dati
- ✅ Jail deaths in America - The U.S. government does not release jail by jail mortality data,
- ✅ Japan
- ✅ Laval, QC, Canada
- ✅ Lexington, KY
- ✅ London Datastore, UK
- ❓ London, ON, Canada
- ✅ Los Angeles Open Data
- ✅ Luxembourg - Luxembourgish Open Data Portal
- ✅ MassGIS, Massachusetts, U.S.
- ✅ Metropolitan Transportation Commission (MTC), California, US
- ✅ Mexico
- ✅ Mississauga, ON, Canada
- ✅ Moldova
- ✅ Moncton, NB, Canada
- ✅ Montreal, QC, Canada
- ✅ Mountain View, California, US (GIS)
- ❓ NYC Open Data
- ✅ NYC betanyc
- ✅ Netherlands
- ✅ New York Department of Sanitation Monthly Tonnage - DSNY Monthly Tonnage Data provides

- ✓ New Zealand
- ? OECD
- ? Oakland, California, US
- ✓ Oklahoma
- ✓ Open Data for Africa
- ✓ Open Government Data (OGD) Platform India
- ✓ OpenDataSoft's list of 1,600 open data
- ✓ Oregon
- ✓ Ottawa, ON, Canada
- ✓ Palo Alto, California, US
- ✓ OpenDataPhilly - OpenDataPhilly is a catalog of open data in the Philadelphia region. In
- ✓ Portland, Oregon
- ✓ Portugal - Pordata organization
- ? Puerto Rico Government
- ? Quebec City, QC, Canada
- ✓ Quebec Province of Canada
- ✓ Regina SK, Canada
- ✓ Rio de Janeiro, Brazil
- ✓ Romania
- ? Russia
- ✓ San Diego, CA
- ✓ San Antonio, TX - Community Information Now - CI:Now is a nonprofit serving Bexar (San
- ✓ San Francisco Data sets
- ✓ San Jose, California, US
- ✓ San Mateo County, California, US
- ? Saskatchewan, Province of Canada
- ✓ Seattle
- ✓ Singapore Government Data
- ✓ South Africa Trade Statistics
- ✓ South Africa
- ✓ State of Utah, US

- ✓ Switzerland
- ✓ Taiwan gov
- ✓ Taiwan
- ✓ Tel-Aviv Open Data
- ✓ Texas Open Data
- ✓ The World Bank
- ✓ Toronto, ON, Canada
- ✓ Tunisia
- ✓ U.K. Government Data
- ✓ U.S. American Community Survey
- ✓ U.S. CDC Public Health datasets
- ✓ U.S. Census Bureau
- ✓ U.S. Department of Housing and Urban Development (HUD)
- ? U.S. Federal Government Agencies
- ✓ U.S. Federal Government Data Catalog
- ✓ U.S. Food and Drug Administration (FDA)
- ✓ U.S. National Center for Education Statistics (NCES)
- ✓ U.S. Open Government
- ✓ UK 2011 Census Open Atlas Project
- ✓ US Counties - This is a repository of various data, broken down by US county. While most of
- ✓ U.S. Patent and Trademark Office (USPTO) Bulk Data Products
- ⚠ Uganda Bureau of Statistics
- ✓ Ukraine
- ✓ United Nations
- ✓ Uruguay
- ✓ Valley Transportation Authority (VTA), California, US
- ⚠ Vancouver, BC Open Data Catalog
- ✓ Victoria, BC, Canada

- ⚠ Vienna, Austria
- ? Statistics from the General Statistics Office of Vietnam - Data in different categories are
- ✓ U.S. Congressional Research Service (CRS) Reports

Healthcare

- ⚠ AWS COVID-19 Datasets - We're working with organizations who make COVID-19-related data
- ✓ COVID-19 Case Surveillance Public Use Data - The COVID-19 case surveillance system database
- ✓ Covid-19 non-processed data of Ecuador - It's a project which provides non-processed datasets
- ⚠ 2019 Novel Coronavirus COVID-19 Data Repository by Johns Hopkins CSSE - This is the data
- ✓ Coronavirus (Covid-19) Data in the United States - The New York Times is releasing a series
- ? COVID-19 Reported Patient Impact and Hospital Capacity by Facility - The following dataset
- ✓ Composition of Foods Raw, Processed, Prepared USDA National Nutrient Database for Standard
- ✓ The COVID Tracking Project - The COVID Tracking Project collects and publishes the most
- ? EHDP Large Health Data Sets
- ✓ GDC - GDC supports several cancer genome programs for CCG, TCGA, TARGET etc.
- ✓ Gapminder World demographic databases
- ✓ MeSH, the vocabulary thesaurus used for indexing articles for PubMed

- ⚠ MeDAL - A large medical text dataset curated for abbreviation disambiguation - Medical
- ⚠ Medicare Coverage Database (MCD), U.S.
- ✓ Medicare Data Engine of medicare.gov Data
- ? Medicare Data File
- ⚠ Nightingale Open Science
- ✓ Number of Ebola Cases and Deaths in Affected Countries (2014)
- ✓ Open-ODS (structure of the UK NHS)
- ✓ OpenPaymentsData, Healthcare financial relationship data
- ✓ PhysioBank Databases - A large and growing archive of physiological data.
- ✓ The Cancer Imaging Archive (TCIA)
- ✓ The Cancer Genome Atlas project (TCGA)
- ✓ World Health Organization Global Health Observatory
- ✓ Yahoo Knowledge Graph COVID-19 Datasets - The Yahoo Knowledge Graph team at Verizon Media is
- ✓ Informatics for Integrating Biology and the Bedside

ImageProcessing

- ✓ 10k US Adult Faces Database
- ⚠ 2GB of Photos of Cats
- ✓ Audience Unfiltered faces for gender and age classification
- ⚠ Affective Image Classification

- Airborne Object Detection and Tracking - The Airborne Object Tracking (AOT) dataset is a
- Animals with attributes
- ✓ CADDY Underwater Stereo-Vision Dataset of divers' hand gestures - Contains 10K stereo pair
- Cytology Dataset – CCAgT: Images of Cervical Cells with AgNOR Stain Technique - Contains 9339
- Caltech Pedestrian Detection Benchmark
- ✓ Chars74K dataset - Character Recognition in Natural Images (both English and Kannada are available)
- ✓ Cube++ - 4890 raw 18-megapixel images, each containing a SpyderCube color target in their
- Densely Annotated Video Driving Data Set - This data set consists of 28 video sequences of
- Danbooru Tagged Anime Illustration Dataset - A large-scale anime image database with 3.33m+
- ? DukeMTMC Data Set - DukeMTMC aims to accelerate advances in multi-target multi-camera
- ETH Entomological Collection (ETHEC) Fine Grained Butterfly (Lepidoptra) Images
- ✓ Face Recognition Benchmark

- ❓ Flickr: 32 Class Brand Logos
- ✅ GDXray - X-ray images for X-ray testing and Computer Vision
- ✅ HumanEva Dataset - The HumanEva-I dataset contains 7 calibrated video sequences (4 grayscale
- ✅ ImageNet (in WordNet hierarchy)
- ✅ Indoor Scene Recognition
- ✅ International Affective Picture System, UFL
- ✅ KITTI Vision Benchmark Suite
- ✅ Labeled Information Library of Alexandria - Biology and Conservation - Contains over 10
- ✅ MNIST database of handwritten digits, near 1 million examples
- ✅ Multi-View Region of Interest Prediction Dataset for Autonomous Driving - Contains 16 driving
- ✅ Massive Visual Memory Stimuli, MIT
- ✅ Newspaper Navigator - This dataset consists of extracted visual content for 16,358,041
- ✅ Open Images From Google - Pictures with segmentation masks for 2.8 million object instances
- ✅ RuFa - Contains images of text written in one of two Arabic fonts (Ruqaa and Nastaliq
- ✅ SUN database, MIT
- ✅ SVIRO Synthetic Vehicle Interior Rear Seat Occupancy - 25.000 synthetic scenery's across ten
- ❓ Several Shape-from-Silhouette Datasets
- ✅ Stanford Dogs Dataset
- ✅ The Action Similarity Labeling (ASLAN) Challenge
- ✅ The Oxford-IIIT Pet Dataset
- ✅ Violent-Flows - Crowd Violence / Non-violence Database and benchmark
- ❓ Visual genome
- ✅ YouTube Faces Database

MachineLearning

- ✓ All-Age-Faces Dataset - Contains 13'322 Asian face images distributed across all ages (from 2
- ✓ Audi Autonomous Driving Dataset - We have published the Audi Autonomous Driving Dataset
- ✓ B3FD - Facial age (and gender) estimation dataset with 375k images - The B3FD dataset is a
- ✓ Context-aware data sets from five domains
- ✓ Delve Datasets for classification and regression
- ✓ Discogs Monthly Data
- ✓ Fluorescent Neuronal Cells - By releasing this dataset, we aim at providing a new testbed for
- ✓ Free Music Archive
- ✓ IMDb Database
- ✓ Iranis - A Large-scale Dataset of Farsi/Arabic License Plate Characters
- ✓ Keel Repository for classification, regression and time series
- ✓ LLVIP - This dataset contains 30976 images, or 15488 pairs, most of which were taken at very
- ✓ Labeled Faces in the Wild (LFW)
- ✓ Lending Club Loan Data
- ? Machine Learning Data Set Repository
- ✓ Million Song Dataset
- ✓ More Song Datasets
- ✓ MovieLens Data Sets
- ✓ New Yorker caption contest ratings
- ? RDataMining - "R and Data Mining" ebook data
- ? Registered Meteorites on Earth
- ✓ Restaurants Health Score Data in San Francisco
- ✓ TikTok Dataset - More than 300 dance videos that capture a single person performing dance
- ✓ UCI Machine Learning Repository
- ✓ Yahoo! Ratings and Classification Data
- ✓ YouTube-BoundingBoxes
- ✓ Youtube 8m
- ✓ eBay Online Auctions (2012)

Museums

- ❓ Canada Science and Technology Museums Corporation's Open Data
- ✅ Cooper-Hewitt's Collection Database
- ✅ Metropolitan Museum of Art Collection API
- ✅ Minneapolis Institute of Arts metadata
- ✅ Natural History Museum (London) Data Portal
- ✅ Rijksmuseum Historical Art Collection
- ✅ Tate Collection metadata
- ✅ The Getty vocabularies

NaturalLanguage

- ✅ Automatic Keyphrase Extraction
- ❓ The Big Bad NLP Database
- ✅ Blizzard Challenge Speech - The speech + text data comes from professional audiobooks
- ✅ Blogger Corpus
- ✅ CLiPS Stylometry Investigation Corpus
- ✅ ClueWeb09 FACC
- ✅ ClueWeb12 FACC
- ✅ DBpedia - Structured data from Wikipedia
- ✅ Dirty Words - With millions of images in our library and billions of user-submitted keywords,
- ❓ Flickr Personal Taxonomies
- ❓ Freebase of people, places, and things
- ✅ German Political Speeches Corpus - Collection of political speeches from the German
- ✅ Google Books Ngrams (2.2TB)
- ✅ Google MC-AFP - Generated based on the public available Gigaword dataset using Paragraph Vectors
- ❓ Google Web 5gram (1TB, 2006)
- ❓ Gutenberg eBooks List

- ? Hansards text chunks of Canadian Parliament
- ✓ LJ Speech - Speech dataset consisting of 13,100 short audio clips of a single speaker reading
- ? M-AILabs Speech - The M-AILABS Speech Dataset is the first large dataset that we are
- ✓ Microsoft MAchine Reading COmprehension Dataset (or MS MARCO)
- ✓ Machine Comprehension Test (MCTest) of text from Microsoft Research
- ✓ Machine Translation of European languages
- ? Making Sense of Microposts 2013 - Concept Extraction
- ✓ Making Sense of Microposts 2016 - Named Entity rEcognition and Linking
- ✓ Multi-Domain Sentiment Dataset (version 2.0)
- ✓ No Language Left Behind (NLLB - 200vo) - Dataset based on Meta's metadata for mined bitext.
- ✓ Noisy speech database for training speech enhancement algorithms and TTS models - Clean and
- ? Open Multilingual Wordnet
- ✓ POS/NER/Chunk annotated data
- ? Personae Corpus
- ? SMS Spam Collection in English
- ✓ SaudiNewsNet Collection of Saudi Newspaper Articles (Arabic, 30K articles)
- ✓ Stanford Question Answering Dataset (SQuAD)
- ✓ USENET postings corpus of 2005~2011
- ✓ Universal Dependencies
- ? Webhose - News/Blogs in multiple languages
- ✓ Wikidata - Wikipedia databases
- ✓ Wikipedia Links data - 40 Million Entities in Context
- ✓ WordNet databases and tools
- ✓ Wordbank - Open, de-identified database of vocabulary development from 84,138 children and
- ✓ WorldTree Corpus of Explanation Graphs for Elementary Science Questions - a corpus of

Neuroscience

- ✓ Allen Institute Datasets
- ✓ Brain Catalogue
- ? Brainomics
- ? CodeNeuro Datasets
- ✓ Collaborative Research in Computational Neuroscience (CRCNS)
- ✓ FCP-INDI
- ✓ Human Connectome Project
- ✓ NDAR
- ✓ NIMH Data Archive
- ✓ NeuroData
- ? NeuroMorpho - NeuroMorpho.Org is a centrally curated inventory of digitally reconstructed
- ✓ Neuroelectro
- ✓ OASIS
- ✓ OpenNEURO
- ✓ OpenfMRI
- ✓ Study Forrest
- ✓ The Nencki-Symfonia EEG/ERP dataset - A high-density electroencephalography (EEG) dataset

Physics

- ✓ CERN Open Data Portal
- ✓ Crystallography Open Database
- ✓ IceCube - South Pole Neutrino Observatory
- ✓ Ligo Open Science Center (LOSC) - Gravitational wave data from the LIGO Hanford and
- ✓ NASA Exoplanet Archive
- ✓ NSSDC (NASA) data of 550 space spacecraft
- ✓ Quantum simulations of an electron in a two dimensional potential well - The data was
- ? Sloan Digital Sky Survey (SDSS) - Mapping the Universe

ProstateCancer

- ✅ EOPC-DE-Early-Onset-Prostate-Cancer-Germany - Early Onset Prostate Cancer - Germany.
- ✅ GENIE - Data from the Genomics Evidence Neoplasia Information Exchange (GENIE) project of the
- ✅ Genomic-Hallmarks-Prostate-Adenocarcinoma-CPC-GENE - Comprehensive genomic profiling of 477
- ✅ MSK-IMPACT-Clinical-Sequencing-Cohort-MSKCC-Prostate-Cancer - Targeted sequencing of clinical
- ✅ Metastatic-Prostate-Adenocarcinoma-MCTP - Comprehensive profiling of 61 prostate cancer
- ✅ Metastatic-Prostate-Cancer-SU2CPCF-Dream-Team - Comprehensive analysis of 150 metastatic
- ✅ NPCR-2001-2015 - Database from CDC's National Program of Cancer Registries (NPCR). The
- ✅ NPCR-2005-2015 - Database from CDC's National Program of Cancer Registries (NPCR). The
- ✅ NaF-Prostate - NaF Prostate is a collection of F-18 NaF positron emission tomography/computed
- ✅ Neuroendocrine-Prostate-Cancer - Whole exome and RNA Seq data of castration resistant
- ✅ PLCO-Prostate-Diagnostic-Procedures - The Prostate Diagnostic Procedures dataset (95,837
- ✅ PLCO-Prostate-Medical-Complications - The Prostate Medical Complications dataset (3,350
- ✅ PLCO-Prostate-Screening-Abnormalities - The Prostate Screening Abnormalities dataset (10,527
- ✅ PLCO-Prostate-Screening - The Prostate Screening dataset (177,315 records, 35,875 subjects,
- ✅ PLCO-Prostate-Treatments - The Prostate Treatments dataset (13,409 records, 7,614 subjects,
- ✅ PLCO-Prostate - The Prostate dataset is a comprehensive dataset that contains nearly all the
- ✅ PRAD-CA-Prostate-Adenocarcinoma-Canada - Prostate Adenocarcinoma - Canada. Collected by the

- ✅ PRAD-FR-Prostate-Adenocarcinoma-France - Prostate Adenocarcinoma - France. Collected by ten
- ✅ PRAD-UK-Prostate-Adenocarcinoma-United-Kingdom - Prostate Adenocarcinoma - United Kingdom.
- ❓ PROSTATEx-Challenge - Retrospective set of prostate MR studies. All studies included
- ✅ Prostate-3T - The Prostate-3T project provided imaging data to TCIA as part of an ISBI
- ✅ Prostate-Adenocarcinoma-Broad-Cornell-2012 - Comprehensive profiling of 112 prostate cancer
- ✅ Prostate-Adenocarcinoma-Broad-Cornell-2013 - Comprehensive profiling of 57 prostate cancer
- ✅ Prostate-Adenocarcinoma-CNA-study-MSKCC - Copy-number profiling of 103 primary prostate
- ✅ Prostate-Adenocarcinoma-Fred-Hutchinson-CRC - Comprehensive profiling of prostate cancer
- ✅ Prostate Adenocarcinoma (MSKCC/DFCI) - Whole Exome Sequencing of 1013 prostate cancer samples.
- ✅ Prostate-Adenocarcinoma-MSKCC - MSKCC Prostate Oncogenome Project. 181 primary, 37 metastatic
- ✅ Prostate-Adenocarcinoma-Organoids-MSKCC - Exome profiling of prostate cancer samples and
- ✅ Prostate-Adenocarcinoma-Sun-Lab - Whole-genome and Transcriptome Sequencing of 65 Prostate
- ✅ Prostate-Adenocarcinoma-TCGA-PanCancer-Atlas - Comprehensive TCGA PanCanAtlas data from 11k
- ✅ Prostate-Adenocarcinoma-TCGA - Integrated profiling of 333 primary prostate adenocarcinoma samples.
- ✅ Prostate-Diagnosis - PCa T1- and T2-weighted magnetic resonance images (MRIs) were acquired
- ❓ Prostate-Fused-MRI-Pathology - The Prostate Fused-MRI-Pathology collection is a combination
- ✅ Prostate-MRI - The Prostate-MRI collection of prostate Magnetic Resonance Images (MRIs) was
- ✅ Prostate-R - The R package 'ElemStatLearn' contains a prostate cancer dataset from Stamey et

- ✅ QIN-PROSTATE-Repeatability - The QIN-PROSTATE-Repeatability dataset is a dataset with
- ✅ QIN-PROSTATE - The QIN PROSTATE collection of the Quantitative Imaging Network (QIN) contains
- ✅ SEER-YR1973_2015.SEER9 - The SEER November 2017 Research Data files from nine SEER registries
- ✅ SEER-YR1992_2015.SJ_LA_RG_AK - The SEER November 2017 Research Data files from the San Jose-
- ✅ SEER-YR2000_2015.CA_KY_LO_NJ_GA - The SEER November 2017 Research Data files from the Greater
- ✅ SEER-YR2000_2015.CA_KY_LO_NJ_GA - The July - December 2005 diagnoses for Louisiana from their
- ✅ TCGA-PRAD-US - TCGA Prostate Adenocarcinoma (499 samples).

Psychology+Cognition

- ❓ OSU Cognitive Modeling Repository Datasets
- ✅ Open Cognitive Science Data - Pubicly available behavioral datasets from across cognitive

PublicDomains

- ✅ Ably Open Realtime Data
- ✅ Amazon
- ✅ Archive.org Datasets
- ✅ Archive-it from Internet Archive
- ✅ CMU JASA data archive
- ✅ CMU StatLab collections
- ✅ Data.World
- ❓ Data360
- ✅ Enigma Public
- ✅ Google
- ❓ Grand Comics Database - The Grand Comics Database (GCD) is a nonprofit, internet-based

- ? Infochimps
- ✓ KDNuggets Data Collections
- ? Microsoft Azure Data Market Free DataSets
- ✓ Microsoft Data Science for Research
- ✓ Microsoft Research Open Data
- ✓ Open Library Data Dumps
- ✓ Reddit Datasets
- ? RevolutionAnalytics Collection
- ✓ Sample R data sets
- ✓ Stack Overflow Annual Developer Survey - Annual developer surverys full data sets from 2011
- ✓ StatSci.org
- ✓ Stats4Stem R data sets (archived)
- ? The Washington Post List
- ✓ UCLA SOCR data collection
- ✓ UFO Reports
- ? Wikileaks 911 pager intercepts
- ✓ Yahoo Webscope

SearchEngines

- ✓ Academic Torrents of data sharing from UMB
- ? Base dos Dados - Data Basis: Open Data Repository for Brazil
- ✓ Datahub.io
- ✓ Domains Project - Sorted list of Internet domains
- ✓ Harvard Dataverse Network of scientific data
- ✓ ICPSR (UMICH)
- ✓ Institute of Education Sciences
- ✓ National Technical Reports Library
- ✓ Open Data Certificates (beta)
- ✓ OpenDataNetwork - A search engine of all Socrata powered data portals
- ✓ Statista.com - statistics and Studies
- ✓ Zenodo - An open dependable home for the long-tail of science

SocialNetworks

- ✅ 2021 Portuguese Elections Twitter Dataset - 57M+ tweets, 1M+ users - This dataset contains
- ✅ 72 hours #gamergate Twitter Scrape
- ✅ CMU Enron Email of 150 users
- ✅ Cheng-Caverlee-Lee September 2009 - January 2010 Twitter Scrape
- ✅ China Biographical Database - The China Biographical Database is a freely accessible
- ✅ Clubhouse Dataset
- ✅ A Twitter Dataset of 40+ million tweets related to COVID-19 - Due to the relevance of the
- ✅ 43k+ Donald Trump Twitter Screenshots - This archive contains screenshots of 43,475 Donald
- ✅ EDRM Enron EMail of 151 users, hosted on S3
- ✅ Facebook Data Scrape (2005)
- ✅ Facebook Social Connectedness Index - We use an anonymized snapshot of all active Facebook
- ✅ Facebook Social Networks from LAW (since 2007)
- ✅ Foursquare from UMN/Sarwat (2013)
- ✅ GitHub Collaboration Archive
- ✅ Google Scholar citation relations
- ✅ High-Resolution Contact Networks from Wearable Sensors
- ✅ Indie Map: social graph and crawl of top IndieWeb sites
- ✅ Mobile Social Networks from UMASS
- ✅ Network Twitter Data
- ❓ Reddit Comments
- ✅ Skytrax' Air Travel Reviews Dataset
- ✅ Social Twitter Data
- ❓ SourceForge.net Research Data
- ✅ The Reddit COVID dataset - This dataset attempts to capture the full extent of COVID-19
- ✅ Twitch Top Streamer's Data
- ✅ Twitter Data for Online Reputation Management

- ✓ Twitter Data for Sentiment Analysis
- ✓ Twitter Graph of entire Twitter site
- ? Twitter Scrape Calufa May 2011
- ✓ UNIMI/LAW Social Network Datasets
- ✓ United States Congress Twitter Data - Daily datasets with tweets of 1100+ accounts associated
- ✓ Yahoo! Graph and Social Data
- ✓ Youtube Video Social Graph in 2007,2008

SocialSciences

- ✓ ACLED (Armed Conflict Location & Event Data Project)
- ✓ Authoritarian Ruling Elites Database - The Authoritarian Ruling Elites Database (ARED) is a
- ✓ Canadian Legal Information Institute
- ? Center for Systemic Peace Datasets - Conflict Trends, Polities, State Fragility, etc
- ✓ Correlates of War Project
- ✓ Cryptome Conspiracy Theory Items
- ? Datacards
- ✓ European Social Survey
- ✓ FBI Hate Crime 2013 - aggregated data
- ? Fragile States Index
- ✓ GDELT Global Events Database
- ✓ General Social Survey (GSS) since 1972
- ✓ German Social Survey
- ✓ Global Religious Futures Project
- ✓ Gun Violence Data - A comprehensive, accessible database that contains records of over 260k
- ✓ Humanitarian Data Exchange
- ✓ INFORM Index for Risk Management
- ✓ Institute for Demographic Studies
- ✓ International Networks Archive

- ✅ International Social Survey Program ISSP
- ✅ International Studies Compendium Project
- ✅ James McGuire Cross National Data
- ✅ MIT Reality Mining Dataset
- ❓ MacroData Guide by Norsk samfunnsvitenskapelig datatjeneste
- ✅ Mass Mobilization Data Project - The Mass Mobilization (MM) data are an effort to understand
- ✅ Microsoft Academic Knowledge Graph - The Microsoft Academic Knowledge Graph is a large RDF
- ✅ Minnesota Population Center
- ✅ Notre Dame Global Adaptation Index (ND-GAIN)
- ✅ Open Crime and Policing Data in England, Wales and Northern Ireland
- ✅ OpenSanctions - A global database of persons and companies of political, criminal, or
- ✅ Paul Hensel General International Data Page
- ✅ PewResearch Internet Survey Project
- ✅ PewResearch Society Data Collection
- ❓ Political Polarity Data
- ✅ StackExchange Data Explorer
- ❓ Terrorism Research and Analysis Consortium
- ❓ Texas Inmates Executed Since 1984
- ✅ Titanic Survival Data Set
- ✅ UCB's Archive of Social Science Data (D-Lab)
- ✅ UCLA Social Sciences Data Archive
- ❓ UN Civil Society Database
- ✅ UPJOHN for Labor Employment Research
- ✅ Universities Worldwide
- ❓ Uppsala Conflict Data Program
- ✅ World Bank Open Data
- ✅ World Inequality Database - The World Inequality Database (WID.world) aims to provide open
- ❓ WorldPop project - Worldwide human population distributions

Software

- ✅ FLOSSmole data about free, libre, and open source software development
- ❓ GHTorrent - Scalable, queryable, offline mirror of data offered through the GitHub REST API.
- ✅ Libraries.io Open Source Repository and Dependency Metadata
- ✅ Public Git Archive - a Big Code dataset for all – dataset of 182,014 top-bookmarked Git
- ✅ Code duplicates - 2k Java file and 600 Java function pairs labeled as similar or different by
- ✅ Commit messages - 1.3 billion GitHub commit messages till March 2019
- ✅ Pull Request review comments - 25.3 million GitHub PR review comments since January 2015 till
- ✅ Source Code Identifiers - 41.7 million distinct splittable identifiers collected from 182,014

Sports

- ✅ American Ninja Warrior Obstacles - Contains every obstacle in the history of American Ninja
- ❓ Betfair Historical Exchange Data
- ✅ Cricsheet Matches (cricket)
- ✅ Equity in Athletics - The Equity in Athletics Data Analysis Cutting Tool is brought to you by
- ✅ Ergast Formula 1, from 1950 up to date (API)
- ✅ Football/Soccer resources (data and APIs)
- ❓ Lahman's Baseball Database
- ✅ NFL play-by-play data - NFL play-by-play data sourced from:
- ✅ Pinhooker: Thoroughbred Bloodstock Sale Data
- ✅ Pro Kabadi season 1 to 7 - Pro Kabadi League is a professional-level Kabaddi league in India.
- ✅ Retrosheet Baseball Statistics
- ✅ Tennis database of rankings, results, and stats for ATP
- ✅ Tennis database of rankings, results, and stats for WTA

- ✓ Transfermarkt Datasets - Clean, structured and automatically updated football (soccer) data
- ✓ USA Soccer Teams and Locations - USA soccer teams and locations. MLS, NWSL, and USL

TimeSeries

- ✓ 3W dataset - To the best of its authors' knowledge, this is the first realistic and public
- ✓ Databanks International Cross National Time Series Data Archive
- ✓ Hard Drive Failure Rates
- ✓ Heart Rate Time Series from MIT
- ✓ Time Series Data Library (TSDL) from MU
- ✓ Turing Change Point Dataset - Contains 42 annotated time series collected for the development
- ✓ UC Riverside Time Series Dataset

Transportation

- ✓ Airlines OD Data 1987-2008
- ✓ Ford GoBike Data (formerly Bay Area Bike Share Data)
- ✓ Bike Share Systems (BSS) collection
- ? Dutch Traffic Information
- ✓ GeoLife GPS Trajectory from Microsoft Research
- ✓ German train system by Deutsche Bahn
- ✓ Hubway Million Rides in MA
- ✓ Montreal BIXI Bike Share
- ✓ NYC Taxi Trip Data 2009-
- ✓ NYC Taxi Trip Data 2013 (FOIA/FOILed)
- ✓ NYC Uber trip data April 2014 to September 2014
- ✓ Open Traffic collection
- ✓ OpenFlights - airport, airline and route data
- ✓ Philadelphia Bike Share Stations (JSON)
- ✓ Plane Crash Database, since 1920

- ? RITA Airline On-Time Performance data
- ? RITA/BTS transport data collection (TranStat)
- ✓ Renfe (Spanish National Railway Network) dataset
- ✓ Toronto Bike Share Stations (JSON and GBFS files)
- ✓ Transport for London (TFL)
- ? Travel Tracker Survey (TTS) for Chicago
- ✓ U.S. Bureau of Transportation Statistics (BTS)
- ✓ U.S. Domestic Flights 1990 to 2009
- ? U.S. Freight Analysis Framework since 2007
- ✓ U.S. National Highway Traffic Safety Administration - Fatalities since 1975 - Contains CSV

eSports

- ✓ CS:GO Competitive Matchmaking Data - In this data set we have data about the CSGO matchmaking
- ✓ FIFA-2021 Complete Player Dataset
- ✓ OpenDota data dump

Complementary Collections

- Data Packaged Core Datasets
- OpenDataMonitor: An overview of available open data resources in Europe
- Quora: Where can I find large datasets open to the public?
- RS.io: 100+ Interesting Data Sets for Statistics
- CVonline: Image Databases
- InnoTrek: Leveraging open data to understand urban lives
- CV Papers: CV Datasets on the web

Credit: https://github.com/awesomedata/awesome-public-datasets

Machine Learning, MLOps & GenerativeAI Roadmap

https://god-level-python.notion.site/Build-a-Strong-Machine-Learning-Portfolio-Personal-Brand-Get-Tons-of-Job-Offers-in-12-Weeks-Live-b3c98407b4ab45819811db081ae9d102?pvs=4

About me

I am Himanshu Ramchandani a Data & Engineering Consultant. I help enterprises utilize big data to build AI-powered products & Mentor professionals to improve their skills in the data field by 1% every day.

the epoch → an AI Newsletter

→ Leverage Data, Products & AI in 3 min.

→ Top 2 AI news & developments.

→ 1 Action Tip from Experts in BigData Analytics, Data Engg & ML.

→ AI Investments.

→ Career & Jobs.

Join the tribe of 20,000+ Entrepreneurs, Tech Leaders, Data Professionals & Devs.

Subscribe to the newsletter here:

https://the-epoch-by-himanshu-ramchandani.beehiiv.com/

Join the Discord Community: https://discord.gg/2Rb9HCpJG