SHORT DESCRIPTION	RESOURCES/DATASETS
Activity classification	https://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring
Air passenger satisfaction	https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction?select=train.csv
Airport density vs. GDP	https://ourairports.com/data/
Alcoholic classification based on EEG data	https://www.kaggle.com/nnair25/Alcoholics
Amazon review polarity prediction	https://www.kaggle.com/kritanjalijain/amazon-reviews?select=train.csv
Amazon videogames recommender system	http://jmcauley.ucsd.edu/data/amazon/
American Sign Language image prediction	https://www.kaggle.com/grassknoted/asl-alphabet
American Sign Language prediction from images	https://www.kaggle.com/datasets/datamunge/sign-language-mnist?select=amer_sign3.png
Anime recommender system	https://www.kaggle.com/azathoth42/myanimelist
Anime recommender system	https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020 https://www.kaggle.com/datasets/nikdavis/steam-store-games
Anime recommender system	https://www.kaggle.com/datasets/svanoo/myanimelist-dataset
Anime recommender system	https://www.kaggle.com/marlesson/myanimelist-dataset-animes-profiles-reviews
Artist/painting classification	https://www.kaggle.com/ikarus777/best-artworks-of-all-time
Asteroid diameter prediction	https://www.kaggle.com/basu369victor/prediction-of-asteroid-diameter?select=Asteroid_Updated.csv
Australian weather forecasting	https://www.kaggle.com/jsphyg/weather-dataset-rattle-package
Auto-insurance claim	https://www.kaggle.com/datasets/bharatnatrayn/auto-insurance-claim
Bank marketing campaign	http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#
Beer rating prediction	https://www.kaggle.com/datasets/rdoume/beerreviews
Beer recommender system	https://www.kaggle.com/code/joonim/beer-reviews-asking-questions/data
Bikesharing ride prediction	https://www.kaggle.com/jackywang529/toronto-bikeshare-data
Bitcoin price increase/decrease from tweets	https://www.kaggle.com/alaix14/bitcoin-tweets-20160101-to-20190329
Bitcoin price prediction	https://query1.finance.yahoo.com/v7/finance/download/BTC-USD?period1=1410998400&period2=1629763200&interval=1d&events=history&includeAdjustedClose=true
Bitcoin price prediction	https://www.kaggle.com/mczielinski/bitcoin-historical-data
Blockchain graph analysis	Collected on-premise
Blog post recommender system	https://www.kaggle.com/vatsalparsaniya/post-pecommendation
Book classification	Collected on-premise (goodreads.com)
Book classification	https://github.com/pgcorpus/gutenberg
Book recommender system	https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/reviews?authuser=0#h.p_bfIevyGdJprA
Book recommender system	https://www.kaggle.com/jirakst/bookcrossing
Book recommender system	https://www.kaggle.com/zygmunt/goodbooks-10k
Botnet detection	http://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT
Buzz prediction in Twitter and in Tom's Hardware	https://archive.ics.uci.edu/ml/datasets/Buzz+in+social+media+#

SHORT DESCRIPTION	RESOURCES/DATASETS
Cerebral strokes prediction	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/44RCPZ
Chat sentiment analyzer	Collected on-premise
Chess game prediction	https://www.kaggle.com/milesh1/35-million-chess-games https://www.kaggle.com/ethanmai/chess-moves
Chess game prediction with LSTM	https://www.kaggle.com/milesh1/35-million-chess-games
Chess match prediction	https://database.lichess.org/
Chess move prediction	https://www.kaggle.com/datasnaek/chess
Chess player clustering	https://www.kaggle.com/datasets/datasnaek/chess
Clustering of research papers on COVID-19	https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge
Code snippet classification	https://www.kaggle.com/simiotic/github-code-snippets
Comparison of recommender systems	https://grouplens.org/datasets/ movielens/
Constrastive learning for 3D point clouds	https://shapenet.org/
Crowd detection from images	http://share.crowdbenchmark.com:2443/home/NWPU-Crowd_Dataset
Customer churn prediction	https://udacity-dsnd.s3.amazonaws.com/sparkify/sparkify_event_data.json
Dark matter particle detection	https://www.kaggle.com/stephenmugisha/particle-collisions https://www.kaggle.com/lavanyashukla01/particle-identification
DDoS attack prediction	https://www.kaggle.com/datasets/devendra416/ddos-datasets
DDoS attack prediction	https://www.kaggle.com/devendra416/ddos-datasets
Drought affection	https://www.gapminder.org/data/ https://github.com/open-numbers/ddf--gapminder--systema_globalis https://github.com/open-numbers/ddf--gapminder--fasttrack
E-commerce purchase prediction	https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store
Eplieptic seizure classification from EEG	https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition
EU ICT industry value prediction	https://data.europa.eu/data/datasets/26d3d1c0-cf84-4f1f-86ad-4ff28c8f7a77?locale=en
Expedia hotel recommendation	https://www.kaggle.com/c/expedia-hotel-recommendations/data
Face mask detection	https://www.kaggle.com/omkargurav/face-mask-dataset
Fake audio detection	http://bil.eecs.yorku.ca/datasets/
Fake news detection	https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset
Fake news detection	https://www.kaggle.com/competitions/fake-news/data
Fake news detection	https://www.kaggle.com/datasets/stevenpeutz/misinformation-fake-news-text-dataset-79k
Fashion product image classification	https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset
Federated learning electric motor temperature	https://www.kaggle.com/wkirgsn/electric-motor-temperature
Flight cancellation prediction	http://stat-computing.org/dataexpo/2009/the-data.html
Flight time prediction	https://www.kaggle.com/datasets/robikscube/flight-delay-dataset-20182022

SHORT DESCRIPTION	RESOURCES/DATASETS
Fog level estimation from images	https://pixabay.com/images/search/foggy%20road/ https://unsplash.com/s/photos/road https://github.com/udacity/self-driving-car/tree/master/datasets
Food image classification	https://www.kaggle.com/datasets/kmader/food41
Football match prediction	https://www.kaggle.com/datasets/stefanoleone992/fifa-21-complete-player-dataset
Football match prediction	www.football-data.co.uk
Football match result prediction + betting	https://www.kaggle.com/datasets/hugomathien/soccer
Football player price prediction	https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset?select=players_18.csv
Forex market forecasting	Collected on-premise (Oanda broker API)
Fraudulent transaction detection	https://www.kaggle.com/c/ieee-fraud-detection
Fraudulent transaction detection	https://www.kaggle.com/competitions/ieee-fraud-detection/data
Fruit image classification	https://github.com/Horea94/Fruit-Images-Dataset
GitHub recommender system	Collected on-premise
GPlay Store sentiment review	Collected on-premise
Health insurance prediction	https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction
Higgs boson particle detection	https://archive.ics.uci.edu/ml/datasets/HIGGS
Home credit eligibility prediction	https://www.kaggle.com/c/home-credit-default-risk/data
Hospital re-admission rate prediction	http://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008
Hotel booking prediction	https://www.kaggle.com/jessemostipak/hotel-booking-demand
Hotel recommender system	https://www.kaggle.com/competitions/expedia-hotel-recommendations/data
House price class prediction	https://www.kaggle.com/hm-land-registry/uk-housing-prices-paid
HuffPost news classification	https://www.kaggle.com/rmisra/news-category-dataset
Human activity prediction from movements	https://www.kaggle.com/gaurav2022/mobile-health
Human age detection from pictures	https://susanqq.github.io/UTKFace/
ID document classification from images	https://www.cs.cmu.edu/~aharley/rvl-cdip/
Image clustering/classification	https://www.kaggle.com/c/landmark-retrieval-2020/data
IMDB movie ratings	https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset
In-vehicle coupon recommendation	http://archive.ics.uci.edu/ml/datasets/in-vehicle+coupon+recommendation
Income prediction	https://archive.ics.uci.edu/ml/datasets/Census+Income
Income prediction	https://www.kaggle.com/lodetomasi1995/income-classification
Instance segmentation	https://www.cityscapes-dataset.com/
Intensive Care Unit prediction	https://www.kaggle.com/competitions/widsdatathon2021
Intrusion detection system	http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
IoT intrusion detection	https://www.kaggle.com/speedwall10/iot-device-network-logs
K-means vs. graph-based clustering	https://openlibrary.org/developers/dumps
League of Legends match prediction	Collected on-premise

SHORT DESCRIPTION	RESOURCES/DATASETS
League of Legends match prediction	https://www.kaggle.com/chuckephron/leagueoflegends
LIGO glitch detection	https://www.kaggle.com/tentotheminus9/gravity-spy-gravitational-waves
Loan prediction	https://www.tandfonline.com/doi/suppl/10.1080/10691898.2018.1434342?scroll=top
Math formula detection	https://www.kaggle.com/datasets/ro101010/math-formula-detection?select=ICDAR2021
Microsoft malware prediction	https://www.kaggle.com/c/microsoft-malware-prediction/data
MIT's GOSSIS (Global Open Source Severity of Illness Score)	https://www.kaggle.com/c/widsdatathon2020
MOOC dropout prediction	http://moocdata.cn/data/user-activity
Movie and TV recommender system from Amazon reviews	https://jmcauley.ucsd.edu/data/amazon/
Movie and TV recommender system from Amazon reviews	https://nijianmo.github.io/amazon/index.html
Movie app	https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset https://www.kaggle.com/datasets/phiitm/movie-posters https://www.kaggle.com/datasets/neha1703/movie-genre-from-its-poster
Movie recommender system	https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset
Movie recommender system	https://www.kaggle.com/rounakbanik/the-movies-dataset
Movie review sentiment analysis	https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews https://github.com/daiquocnguyen/SAR14
MUBI movie recommender system	https://www.kaggle.com/clementmsika/mubi-sqlite-database-for-movie-lovers/tasks?taskId=1150
Multiclass sentiment analysis of Amazon reviews	https://www.kaggle.com/rogate16/amazon-reviews-2018-full-dataset
Netflix prize	https://www.kaggle.com/netflix-inc/netflix-prize-data
Network traffic analyzer	https://www.unb.ca/cic/datasets/nsl.html
New York City Taxi Fare Prediction	https://www.kaggle.com/c/new-york-city-taxi-fare-prediction/overview
New York City Taxi trip time prediction	https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
News popularity prediction	https://www.kaggle.com/datasets/nikhiljohnk/news-popularity-in-multiple-social-media-platforms?select=test_file.csv
Particle collision prediction	http://mlphysics.ics.uci.edu/data/hepjets/
Personality classification from blog posts	https://www.kaggle.com/datasnaek/mbti-type
Plant disease detection from images	https://www.kaggle.com/datasets/vipooooool/new-plant-diseases-dataset
Plant disease detection from images	https://www.kaggle.com/rashikrahmanpritom/plant-disease-recognition-dataset
Potentially hazardous asteroid detection	https://www.kaggle.com/sakhawat18/asteroid-dataset
Predicting New York City's taxi trip duration	https://www.kaggle.com/datasets/elemento/nyc-yellow-taxi-trip-data
Protein classification from images	https://www.kaggle.com/c/human-protein-atlas-image-classification/overview
Psychological trait prediction	https://www.kaggle.com/tunguz/big-five-personality-test?
PUBG match deaths	https://www.kaggle.com/skihikingkevin/pubg-match-deaths
Purchase price prediction on Black Friday	https://www.kaggle.com/sdolezel/black-friday
Real-time object detection	https://waymo.com/open/
Recommender systems for food delivery	https://www.kaggle.com/code/mobasshir/recommendation-engine/data?select=train_full.csv

SHORT DESCRIPTION	RESOURCES/DATASETS
Reddit spoiler vs. non-spoiler classification	Collected on-premise
Review sentiment classification	https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews?select=train.csv
RL-based chess opponent	https://www.kaggle.com/milesh1/35-million-chess-games
Salary prediction	Collected on-premise
San Francisco's optimal police car deployment	https://data.sfgov.org/Public-Safety/Police-Department-Calls-for-Service/hz9m-tj6z
Sarcasm detection in Twitter	Collected on-premise
Semi-supervised image classification	https://cs.stanford.edu/~acoates/stl10/
Sentiment analyzer	https://www.kaggle.com/datasets/visalakshiiyer/twitter-data-brexit?select=TweetDataset_ProBrexit_Jan-Mar2022.csv
Smart meter reading prediction	https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london
Solar irradiance prediction	https://nsrdb.nrel.gov/
Solar irradiance prediction	https://www.kaggle.com/datasets/dronio/SolarEnergy
Song's release prediction	https://archive.ics.uci.edu/ml/datasets/YearPredictionMSD#
Sport activity detection	https://sites.google.com/eng.ucsd.edu/fitrec-project/home
Spotify genre classification + clustering	https://www.kaggle.com/yamaerenay/spotify-tracks-dataset-19222021
Spotify music sentiment analysis	https://www.kaggle.com/edenbd/150k-lyrics-labeled-with-spotify-valence https://www.kaggle.com/neisse/scrapped-lyrics-from-6-genres?select=lyrics-data.csv
Spotify popularity song classification	https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs
Spotify recommender system	https://www.kaggle.com/datasets/ektanegi/spotifydata-19212020
Spotify song clustering	https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks http://millionsongdataset.com/
StackOverflow multi-tag prediction	https://www.kaggle.com/c/facebook-recruiting-iii-keyword-extraction/data
StackOverflow topic modeling	https://archive.org/download/stackexchange
Stellar classification	https://www.kaggle.com/datasets/fedesoriano/stellar-classification-dataset-sdss17
Stock market forecasting	Collected on-premise
Stock market forecasting	https://www.kaggle.com/aaron7sun/stocknews https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs
Stock market forecasting	https://www.kaggle.com/jacksoncrow/stock-market-dataset
Stock market forecasting	https://www.kaggle.com/jerryhans/key-statistics-yahoo-finance-stocks-of-2003-2013 https://www.kaggle.com/jerryhans/key-statistics-yahoo-finance-stocks-of-2003-2013
Stock market forecasting	https://www.kaggle.com/qks1lver/amex-nyse-nasdaq-stock-histories?select=fh_5yrs.csv
Stream frequency prediction (https://openreview.net/pdf?id=r1lohoCq\	https://www.caida.org/data/passive/passive_dataset_request.xml
Tennis match prediction	https://www.kaggle.com/sijovm/atpdata
Time series weather forecasting	https://data.london.gov.uk/dataset/photovoltaic--pv--solar-panel-energy-generation-data
Traffic-related predictions	https://dati.comune.roma.it
Tree cover type prediction on forests	https://archive.ics.uci.edu/ml/datasets/Covertype

SHORT DESCRIPTION	RESOURCES/DATASETS
Twitter bot account detection	https://www.kaggle.com/davidmartngutirrez/twitter-bots-accounts
Twitter hate speech recognition	https://www.kaggle.com/datasets/andre112/measuring-hate-speech
Twitter real-time analytics framework	Collected on-premise
Twitter sentiment analysis	https://www.kaggle.com/kazanova/sentiment140
Twitter sentiment analyzer for movies	http://help.sentiment140.com/for-students http://ai.stanford.edu/~amaas/data/sentiment/
UK used car price prediction	https://www.kaggle.com/adityadesai13/used-car-dataset-ford-and-mercedes
URL classification (phishing/non-phishing)	https://www.kaggle.com/datasets/shashwatwork/web-page-phishing-detection-dataset
Used car price prediction	https://www.kaggle.com/datasets/245df9d02ee0cdff29a4cb6f099efb4e9f4bf69a1ff652838dde679450491a7a?resource=download
User-Movie rating prediction	https://www.kaggle.com/competitions/21422netflix
Videogame recommender system	Collected on-premise
Videogame recommender system	https://cseweb.ucsd.edu/~jmcauley/datasets.html#steam_data
Videogame recommender system	https://www.kaggle.com/datasets/trentenberam/metacritic-games-all-time?select=metacritic_games.csv
Walmart sales forecasting	https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data
Weather forecasting	https://www.kaggle.com/datasets/keshan/the-shifts-weather-prediction-dataset
Weather forecasting	https://www.kaggle.com/selfishgene/historical-hourly-weather-data
Website phishing detection	https://www.kaggle.com/aman9d/phishing-data
Wine variety prediction	https://www.kaggle.com/zynicide/wine-reviews?select=winemag-data_first150k.csv
Yelp review sentiment prediction	https://www.kaggle.com/shikhar42/yelps-dataset