1) Iris Flowers Classification ML Project– Learn about Supervised Machine Learning Algorithms

Iris flowers dataset is one of the best dataset in classification literature. The classification of iris flowers machine learning project is often referred to as the “Hello World” of machine learning. The dataset has numeric attributes and beginners need to figure out on how to load and handle data. The iris dataset is small which easily fits into the memory and does not require any special transformations or scaling to begin with.

Iris Dataset can be downloaded from UCI ML Repository – Download Iris Flowers Dataset

The goal of this machine learning project is to classify the flowers into among the three species – virginica, setosa, or versicolor based on length and width of petals and sepals.

2) BigMart Sales Prediction ML Project – Learn about Unsupervised Machine Learning Algorithms

BigMart sales dataset consists of 2013 sales data for 1559 products across 10 different outlets in different cities. The goal of the BigMart sales prediction ML project is to build a regression model to predict the sales of each of 1559 products for the following year in each of the 10 different BigMart outlets. The BigMart sales dataset also consists of certain attributes for each product and store. This model helps BigMart understand the properties of products and stores that play an important role in increasing their overall sales.

You can access the complete solution to this ML Project Here – BigMart Sales Prediction Machine Learning Project Solution

3) Social Media Sentiment Analysis using Twitter Dataset

Social media platforms like Twitter, Facebook, YouTube, Reddit generate huge amounts of big data that can be mined in various ways to understand trends, public sentiments and opinions. Social media data today has become relevant for branding, marketing, and business as a whole. A sentiment analyser learns about various sentiments behind a “content piece” (could be IM, email, tweet or any other social media post) through machine learning and predicts the same using AI.Twitter data is considered as a definitive entry point for beginners to practice sentiment analysis machine learning problems. Using Twitter dataset, one can get captivating blend of tweet contents and other related metadata such as hashtags, retweets, location, users and more which pave way for insightful analysis. Twitter dataset consists of 31,962 tweets and is 3MB in size. Using Twitter data you can find out what the world is saying about a topic whether it is movies, sentiments about US elections or any other trending topic like predicting who would win the FIFA world cup 2018.Working with the twitter dataset will help you understand the challenges associated with social media data mining and also learn about classifiers in depth. The foremost problem that you can start working on as a beginner is to build a model to classify tweets as positive or negative.

4) Sales Forecasting using Walmart Dataset

Walmart dataset has sales data for 98 products across 45 outlets. The dataset contains sales per store, per department on weekly basis. The goal of this machine learning project is to forecast sales for each department in each outlet to help them make better data driven decisions for channel optimization and inventory planning. The challenging aspect of working with Walmart dataset is that it contains selected markdown events which affect sales and should be taken into consideration.

5) Learn to build Recommender Systems with Movielens Dataset

From Netflix to Hulu, the need to build an efficient movie recommender system has gain importance over time with increasing demand from modern consumers for customized content. One of the most popular dataset available on the web for beginners to learn building recommender systems is the Movielens Dataset which contains approximately 1,000,209 movie ratings of 3,900 movies made by 6,040 Movielens users. You can get started working with this dataset by building a world-cloud visualization of movie titles to build a movie recommender system.

6) Stock Prices Predictor

This is another interesting machine learning project idea for data scientists/machine learning engineers working or planning to work with finance domain. Stock prices predictor is a system that learns about the performance of a company and predicts future stock prices. The challenges associated in working with stock prices data is that it is very granular, and moreover there are different types of data like volatility indices, prices, global macroeconomic indicators, fundamental indicators , and more. One good thing about working with stock market data is that the financial markets have shorter feedback cycles making it easier for data experts to validate their predictions on new data. To begin working with stock market data, you can pick up a simple machine learning problem like predicting 6 month price movements based on fundamental indicators from an organizations' quarterly report.You can download Stock Market datasets from Quandl.com or Quantopian.com.

7) Predicting Wine Quality using Wine Quality Dataset

I love everything that's old,—old friends, old times, old manners, old books, old wine.—Oliver Goldsmith

It's a known fact that older the wine, better the taste. However, there are several factors other than age that go into wine quality certification which include physiochemical tests like alcohol quantity, fixed acidity, volatile acidity, determination of density, pH and more. The main goal of this machine learning project is to build a machine learning model to predict the quality of wines by exploring their various chemical properties. Wine quality dataset consists of 4898 observations with 11 independent and 1 dependent variable.

Get access to the complete solution of this machine learning project here – Wine Quality Prediction in R

8) Boston Housing Price Prediction ML Project

Boston House Prices Dataset consists of prices of houses across different places in Boston. The dataset also consists of information on areas of non-retail business (INDUS), crime rate (CRIM), age of people who own a house (AGE) and several other attributes (the dataset has a total of 14 attributes).Boston Housing dataset can be downloaded from the UCI Machine Learning Repository. The goal of this machine learning project is to predict the selling price of a new home by applying basic machine learning concepts on the housing prices data. This dataset is too small with 506 observations and is considered a good start for machine learning beginners to kick-start their hands-on practice on regression concepts.

9) MNIST Handwritten Digit Classification

Deep learning and neural networks play a vital role in image recognition, automatic text generation, and even self-driving cars. To begin working in these areas, you need to begin with a simple and manageable dataset like MNIST dataset. It is difficult to work with image data over flat relational data and as a beginner we suggest you can pick up and solve the MNIST Handwritten Digit Classification Challenge. MNIST dataset is too small to fit into your PC memory and beginner-friendly. However, the handwritten digit recognition will challenge you.

Make your classic entry into solving image recognition problems by accessing the complete solution here – MNIST Handwritten Digit Classification.

10) Human Activity Recognition using Smartphone Dataset

The smartphone dataset consists of fitness activity recordings of 30 people captured through smartphone enabled with inertial sensors. The goal of this machine learning project is to build a classification model that can precisely identify human fitness activities. Working on this machine learning project will help you understand how to solve multi-classification problems.

Get access to this machine learning projects source code here -Human Activity Recognition