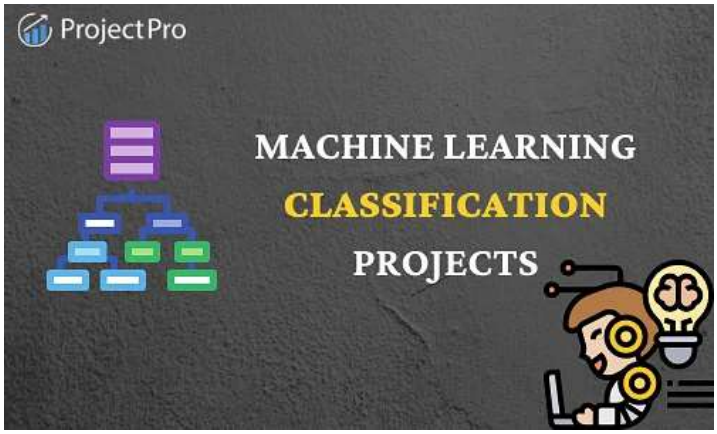
10 Machine Learning Classification Project Ideas for Beginners

Explore solved end-to-end classification project ideas in machine learning to gain hands-on experience and boost your portfolio to get hired.

Last Updated: 28 Mar 2023



Classification is a supervised machine learning problem requiring the model to label or assign a class (from a fixed number of classes) to an example. The familiar problems of classifying email as spam or not spam, predicting the handwritten character, and so on are all examples of machine learning projects on classification.

Classification problems can be broadly classified as binary and multi-class classification, which involves classification into two and more than two classes, respectively, and multi-label classification capable of assigning more than one label for each example. Further, for problems with a large imbalance in the proportion of examples belonging to each class, special considerations need to be made to ensure the accuracy of predictions.

Classes of classification problems aside (Did you see what we did there? ðŸ˜›).

Build a Multi Class Image Classification Model Python using CNN

Downloadable solution code | Explanatory videos | Tech Support

Start Project

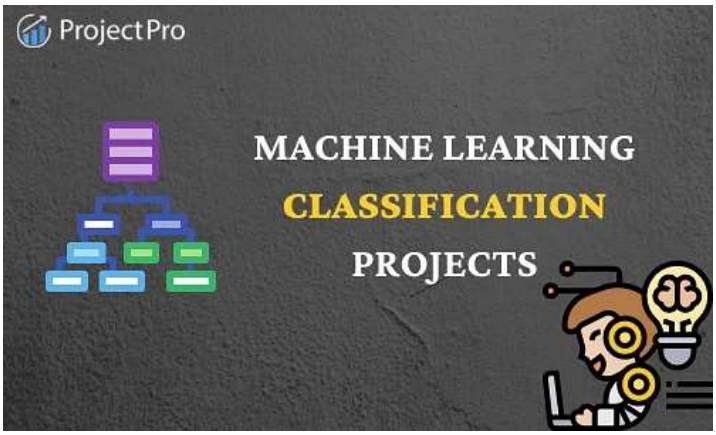
Machine learning algorithms let us make sense of the data, find patterns based on similarity in features/attributes, and help quick decision-making across different industries, including economics, healthcare, e-commerce, retail, etc. As an aspiring machine learning engineer, the best way to accelerate your career and improve your skills is to practice as many hands-on classification projects on machine learning as you can.

NOTE: There is an 'unsupervised version' of classification. This class of problems is, however, more likely to be known as 'clustering.'

Table of Contents

10 Machine Learning Projects on Classification

To understand how machine learning classification algorithms work, one must work on a few innovative machine learning project ideas that implement them. Let's jump on to the classification projects in machine learning and gain hands-on experience in real-time without much ado.



Machine Learning Classification Projects in the Finance Domain

Source: Me.me

View Project | View Project | View Project | View Project

View all New Projects

1. Loan Eligibility Prediction Classification Project

Credit appraisal is conducted in a detailed and systematic manner to gauge a loan applicant's repayment ability. For corporate entities, such an evaluation is a comprehensive process covering the technical and financial elements. A borrower should repay the loan on time without missing deadlines, crucial for a bank's interest income and capital. Thereby, a borrower's repayment behavior directly affects the performance of the bank. However, the appraisal process can be tedious and calls for a degree of automation using machine learning.

You can use the Loan Data Set from Kaggle to build a classification machine learning model for the loan eligibility evaluation use case. You can use a simple classifier model such as the Support Vector Classifier and use the ROC (receiver operating characteristic) curve to evaluate its performance and quantify it using AUC (area under the curve).

Classification Project for Loan Eligibility Prediction

2. Credit Card Fraud Detection Project

Credit fraud involves the illegal use of someone else's credentials to make fraudulent credit card transactions so that the individual is charged for items that they did not purchase. The person whose credit card is mishandled usually ends up with unpaid debt. He/she may be unable to obtain new credit due to a poor credit score (at least until matters are sorted out), which makes early detection of fraudulent transactions by credit card companies crucial.

The publicly available Credit Card Fraud Detection Kaggle dataset, which contains details of the transactions made by European cardholders, can be used to implement this project. The positive class (occurrence of frauds) accounts for only 0.172% of all transactions in the dataset. While you could use any binary classification technique, make sure you take care of the class imbalance. Further, it is recommended to use the area Under the Precision-Recall Curve (AUPRC) to measure the accuracy as confusion matrix accuracy is not meaningful for unbalanced classification.

Credit Card Fraud Detection as a Classification Problem

Machine Learning Classification Projects in the Healthcare Domain

Source: Esmemes

3. Water Quality Prediction using Machine Learning Classification

The first step to effective healthcare is preventing disease. Widespread access to clean water, sanitation, and hygiene are crucial to ensure this. Water quality varies depending on the place and condition of the source of water and the treatment it receives. There could be specific contaminants in water resulting in health issues, ranging from gastrointestinal illness, reproductive problems, and neurological disorders. Infants, children, pregnant women, the elderly, and those with compromised immune systems may be especially susceptible to this.

This classification project involves building a model to evaluate water potability using the following Water Quality dataset. Use the Random Forest Classifier and build a model to assess various properties, including temperature, acidity, turbidity, and hardness, to arrive at the

4. Image-Based Disease Diagnosis

Computed tomography, or CT, is a computerized x-ray imaging procedure used to generate cross-sectional body images. These slices contain more detailed information than conventional x-rays as they can be stacked together to form a three-dimensional image which allows for easier identification and location of abnormalities.

You can use a publicly available Chest CT Scan dataset (https://www.kaggle.com/mohamedhanyyy/chest-ctscan-images) to build a capable neural network of detecting the presence or absence of chest cancer and classifying the images into Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma. You can use a Convolutional neural network (CNN) of your choice to accomplish this.

For Further Reading -

- **Deep Learning for Image Classification with Python for Pneumonia Detection**
- **15 OpenCV Projects Ideas for Beginners to Practice**
- **20 Machine Learning Projects That Will Get You Hired**
- **15 Top Machine Learning Projects for Final Year Students**
- **8 Healthcare Machine Learning Project Ideas for Practice**

Machine Learning Classification Projects in the Entertainment/ Media Domain

Source: Sharecopia

5. Sentiment Analysis Project

Sentiment analysis can help entertainment companies, streaming services, and media platforms to serve their customers or users better by making better recommendations and ensuring a better user experience. With the widespread use of social media, such a sentiment analysis can further help public safety and health, resulting in early recognition of mental health issues and even threats.

Using the Sentimental Analysis for Tweets dataset, you can implement a classifier model to predict whether an individual is depressed or not. One way to accomplish this is to use Term Frequency Inverse Document Frequency (TF-IDF) vectorization followed by a classifier like Logistic Regression or Support Vector Classifier.

Start this Sentiment Analysis Project to Learn More on how to implement sentiment analysis in the real world.

Explore Categories

6. Movie Genre prediction

Source: Reddit

Movie posters are a crucial element in the film industry as they seek to convey the message and feel of the film while piquing the audience's curiosity and attracting them to the theaters. There is more to designing movie posters than just plastering the movie name across a fancy-looking picture. The final Harry Potter movie, for instance, held merely the words "It all ends" and yet managed to draw in scores of fans who flocked the theatres knowing what the movie was named, and at least a few others who just wanted in on the action. Therefore, it is pretty obvious about the impact movie posters have to capture and attract their viewers' attention.

Use the Movie Poster dataset available on Kaggle, which contains movie posters labeled with the movie title, release year, IMDB rating, and genre, train a neural network to predict the movie genre given an unseen example. Keep in mind that this will be a multi-label problem, i.e., each movie should be assigned multiple genre labels like comedy, horror, romance, etc., and you must build the neural network accordingly.

Machine Learning Classification Projects in the Retail/eCommerce Domain

Source: Facebook

7. E-commerce Customer Behavior

Online consumer behavior is the process by which consumers make purchasing decisions on e-commerce platforms. Customer behavior, including how the products are viewed, added to the cart, removed from the cart, and purchased, provides insight into the customers' expectations. While there are many variations between individual shoppers and needs, overall online consumer behavior for expectations like product availability, delivery transparency, and affordable shipping has much in common between individuals. Depending on the ongoing trend of expectations, all the previously mentioned factors influence the decision to purchase and whether the customers will remain loyal once they've made a purchase.

Using eCommerce behavior data from a multi-category store, you can build a model to predict the event type or whether a product added to the cart is purchased using a Random Forest or XGB Classifier.

8. Automated Product Checkout

Automatic checkout (ACO) is one of the critical use cases explored in applying computer vision technology to the retail industry. It is aimed at designing solutions for automatically generating shopping lists from the images of the products to be purchased. Such a solution must include the capability to identify single as well as multi-product images taken by the checkout system.

This classification project aims to build a neural network capable of identifying the instances of various retail products in the images produced by the checkout system (as provided in this dataset: https://www.kaggle.com/diyer22/retail-product-checkout-dataset). Keep in mind that this solution will require object instance-level recognition and multi-label output, as is the case with popular object detection algorithms like YOLOv3.

Machine Learning Classification Projects in the Marketing Domain

Source: Digitalmarketer

9. Customer Segmentation

Customer segmentation involves partitioning the customer base into groups that are similar in specific ways concerning marketing. The factors considered for such a division might include age, gender, interests, and spending habits. The goal of segmenting customers is to maximize the value of each customer to the business by operating with the understanding that every customer is different and thereby allowing marketers to devise marketing strategies that would address each customer in a personalized and effective way. Further, companies can also better understand their customers' preferences and what each customer sub-group finds most valuable to tailor marketing materials more accurately.

To implement this project use case, train a neural network capable of providing a multi-class output with four classes. The Customer Segmentation dataset can be used for this project.

10. Discount Coupons Availed

Promotional techniques like discounts and coupons are widely used to attract new customers and reinforce customer loyalty. Measuring the likelihood of a customer to avail coupons and predicting the redemption behavior are both crucial in assessing the effectiveness of any such marketing campaign. Promotions, including coupon discounts for a product/ range of products, are usually done across various channels, including email and text messages. Owing to the investment made towards such promotions, it is also important to predict whether customers redeem the coupons received across channels and use these predictions to develop more precise and targeted marketing strategies.

This machine-learning classification project aims to predict the coupon redemption status using a binomial classifier using the following publicly available Kaggle dataset: https://www.kaggle.com/meghakanojia/predicting-coupon-redemption. The dataset contains information regarding user demographics, campaign and coupon details, product details, and also previous transactions which should help you build a Decision Tree Classifier.

Machine Learning Classification - A Class Apart!

Due to the widespread use cases associated with classification problems, exploring and understanding the various Classification algorithms is a great way to boost your machine learning expertise. And if it hasn't been stressed enough already (let me just state the obvious again), there is no better way to get better at solving machine learning problems than to solve machine learning problems!

So if you've made it this far through the article, make sure you pat yourself on the back and, for beginners, pick up a project that appeals the most to you and start implementing it right away. By the time you find yourself coming back a couple of times for more projects, your knowledge of classification algorithms will surely be a class apart!

PREVIOUS NEXT

Trending Project Categories

Machine Learning Projects

Data Science Projects

Deep Learning Projects

Big Data Projects

Apache Hadoop Projects

Apache Spark Projects

Show more

Trending Projects

Walmart Sales Forecasting Data Science Project

BigMart Sales Prediction ML Project

Music Recommender System Project

Credit Card Fraud Detection Using Machine Learning

Resume Parser Python Project for Data Science

Time Series Forecasting Projects

Show more

Trending Blogs

Machine Learning Projects for Beginners with Source Code

Data Science Projects for Beginners with Source Code

Big Data Projects for Beginners with Source Code

IoT Projects for Beginners with Source Code

Data Analyst vs Data Scientist

Data Science Interview Questions and Answers

Show more

Trending Recipes

Search for a Value in Pandas DataFrame

Pandas Create New Column based on Multiple Condition

LSTM vs GRU

Plot ROC Curve in Python

Python Upload File to Google Drive

Optimize Logistic Regression Hyper Parameters

Show more

Trending Tutorials

PCA in Machine Learning Tutorial

PySpark Tutorial

Hive Commands Tutorial

MapReduce in Hadoop Tutorial

Apache Hive Tutorial -Tables

Linear Regression Tutorial

Show more

ProjectPro