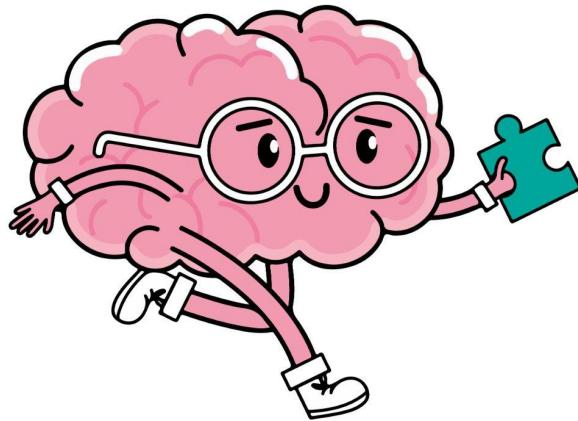


30+ Project Ideas to Showcase Your Machine Learning Skills

Here are 30+ machine learning project ideas ideal for showcasing and honing your ML skills. There's something for everyone, from beginners to advanced levels.



30+ Project Ideas to Showcase Your Machine Learning Skills



Machine learning is a booming field of study as its application plays an important role in basically any industry. Thus, no wonder that currently, the demand for data scientists and machine learning engineers is very high.

However, the high demand for data scientists and machine learning engineers doesn't necessarily mean that it's easy to get a job in these two positions. In fact, the opposite is true. It's very challenging to get a job as a data scientist or machine learning engineer, especially if you don't have a lot of experience beforehand.

As you can imagine, in each data scientist vacancy, most likely, there are hundreds of applicants. This means that you're competing with a lot of people to get that one job. Hence, it's extremely important for you to have something unique that sets you apart

from the other candidates, something unique that can showcase your machine learning skills and impress prospective employees.

Among other things, having a collection of personal projects is one of the best ways to showcase your machine learning skills, and this is what we're going to discuss in this article.

In this article, you will find various personal project ideas that you can implement to showcase your machine learning skills or upskill your knowledge. It doesn't matter whether you're just starting out your journey in learning machine learning or already advanced because we're going to discuss personal projects on three different levels: beginner, intermediate, and advanced.

So without further ado, let's start with project ideas for the beginner level.

Beginner-Level Project Ideas

BEGINNER LEVEL

PROJECT IDEAS



When you're just starting out your journey in machine learning, then the best way to kickstart it is by doing a project with a simple dataset and with the help of an established machine learning library like scikit-learn. With those two things in mind, the best project to start with is by implementing simple [machine learning algorithms](#), like linear regression or simple classification with logistic regression.

Next, after you get a sense of how machine learning in general works, then you can move to slightly advanced machine learning algorithms such as [Decision Tree and Random Forest](#), [Support Vector Machine](#), Naive Bayes, etc.

Machine Learning Project Idea #1: Classifying Iris Flower

Classifying iris flowers would be the best introduction to the world of machine learning classification algorithms if we're just starting out our machine learning journey. To do this project, we will need to use the [Iris Flower dataset](#). This dataset consists of hundreds of flowers in three different iris classes: Setosa, Versicolor, and Virginica. Our task is to build a machine learning algorithm that can classify the class of each flower based on its features: sepal length, sepal width, petal length, and petal width.

This dataset is perfect for beginners because it's clean and ready to use. You can skip the data preprocessing part with this data and feed it to the machine learning algorithm of your choice directly. If you need inspiration on how to do this project step-by-step, [check out this tutorial from Kaggle](#).

Machine Learning Project Idea #2: Predicting the Fate of Titanic Passenger

This project is also very good for us if we're new to machine learning and want to deepen our knowledge not only on classification machine learning algorithms but also on how to conduct proper data preprocessing steps. As the title suggests, the goal of this project is to predict whether a particular passenger survived the Titanic tragedy, given several features such as their sex, age, number of siblings or children, ticket class, etc.

[The Titanic dataset](#) is slightly more challenging than the Iris dataset because before we feed the data into a machine learning algorithm of our choice, we need to do data preprocessing first, such as we need to handle the missing values, doing one-hot encoding, or even doing feature engineering.

If you need a step-by-step example of how to conduct this project, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #3: Predicting Housing Prices in Boston

If you want to get started with machine learning algorithms for regression problems, this project would be the best starting point. The goal of this project is to predict the price of a particular house given several features such as the crime rate of its location, the average number of rooms per dwelling, its tax rate per \$10,000, and so on. In total, there are 14 features.

The dataset that we can use to do this project is [the Boston housing prices](#). In the beginning, we can try to use all of the features available for our regression model and then observe our model's performance using common regression metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). Next, try to calculate the importance of each feature and see if the MSE or RMSE of our model improves when we eliminate some of the features in the dataset.

If you need step-by-step guidance on how to do this project, check out [this tutorial from Kaggle](#).

Intermediate-Level Project Ideas

INTERMEDIATE LEVEL

PROJECT IDEAS



After conducting all of the projects above, we should now be familiar with how to use machine learning for classification and regression tasks. In the intermediate-level projects, we will continue with regression and classification tasks, but with the addition

of more challenges and techniques related to exploratory data analysis, data preprocessing, and feature engineering parts.

In addition to that, we will also cover the unsupervised domain within the machine learning spectrum, such as clustering and anomaly detection. Finally, we will start implementing the basic concept of deep learning for various applications, such as computer vision and Natural Language Processing (NLP).

Machine Learning Project Idea #4: Laptop Price Prediction

Continuing from our last project, we're also dealing with a regression problem in this project, in which we need to build a machine learning model to predict the laptop price given certain features.

The data that we can use for this project comes from Allegro, where our target variable is the laptop's price, and there are a bunch of features that we can use for our machine learning model, such as the number of CPU cores, RAM size, graphic card type, etc. Thus, we have the opportunity to be creative with the data that we have in terms of feature engineering as well as the data preprocessing pipeline.

The dataset and the guidance on how to do this project are available on [StrataScratch here](#).

Machine Learning Project Idea #5: Predicting Movie Rating on Rotten Tomatoes

For this project, we can use the Rotten Tomatoes dataset that consists of a list of movies, their respective rating, and a bunch of features such as movies' release date, their genre, how many people have rated those movies, etc.

The main goal of this project is to build a machine learning model that can predict the Rotten Tomatoes rating of a movie, given the features. The exciting thing about this project is that we can be creative with feature engineering steps, and we can build several classification machine learning models and then compare their performance in the end.

The dataset and the guidance on how to do this project are available on [StrataScratch here](#).

Machine Learning Project Idea #6: Predicting Customer Churn

Customer churn can be described as an event where the customer stopped using a particular company's service. Thus, customer churn is one of the common real-life problems where the application of machine learning would be useful.

In this project, we're going to tackle a data science problem that has been used in the recruitment process at Sony research. There are a lot of features in the data that we're working on, such as the number of days each customer has been using a service, the origin of the customers, the total number of calls made by customers, etc.

Before building a machine learning model for classification, we need to do exploratory data analysis first, and this is a perfect way to showcase our data analytics skills. Next, we can start to build various machine learning classifiers and then compare the pros and cons of each model.

The dataset and the guidance on how to do this project are available on [StrataScratch here](#).

Machine Learning Project Idea #7: Predicting Delivery Duration

When you order food online, you'll most likely get the estimated time of arrival (ETA) of your food. For such a company providing this online service, giving us an accurate ETA would be crucial for customer experience. Thus, creating a machine learning model that is able to predict the delivery duration of an online order would be the goal of this project.

The data that we'll be working on in this project come from DoorDash. Specifically, the dataset consists of a bunch of features such as the number of items in an order, the number of distinct items in an order, the time the order was made, the time the order was delivered to customers, etc.

The dataset and the guidance on how to do this project are available on [StrataScratch here](#).

Machine Learning Project Idea #8: Customer Segmentation for Marketing Campaign

In all of our projects above, we've been working on machine learning projects within a supervised domain, such as classification and regression problems. In this project,

we're going to do a project related to the unsupervised domain in a machine learning spectrum. In unsupervised learning methods, our dataset doesn't come with the true label. This means that we let our machine learning model learn the pattern of our dataset by itself and then make a prediction based on that.

Among all of the unsupervised learning methods, customer segmentation is one of its most common applications in real-life. The dataset that we can do to do this project is [the marketing campaign dataset](#) that consists of the list of purchases made by customers.

With the dataset above, we can do a lot of different methods to cluster the customers based on their purchase history with the help of common machine learning models such as k-means clustering.

If you need step-by-step guidance on how to do this project, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #9: Fraud Detection of Credit Card Transaction

Another common application of unsupervised machine learning methods in real life is anomaly detection, such as credit card fraud detection. The main goal of our project is to build a machine learning model that can detect whether a credit card transaction is normal or fraudulent.

One of the challenges when dealing with anomaly detection is the imbalance dataset, where the number of 'normal transaction' data is far superior to 'fraudulent transaction' data. Thus, in this project, we have the opportunity to play around with the machine learning model's metrics other than accuracy, such as precision, recall, AUC score, or F1 score.

We can also try different preprocessing methods to deal with imbalanced datasets, such as undersampling, oversampling, or applying custom weights to our model before the training process. Then, we can write a summary of the pros and cons of each of the methods mentioned above.

To do this project, we can use [the credit card dataset](#). If you need step-by-step guidance to do this project, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #10: Classifying Handwritten Digits

Now that you have implemented various machine learning algorithms in the previous projects, then it's time for us to get into the concept of deep learning. A common application for deep learning includes computer vision, which we're going to do in this project.

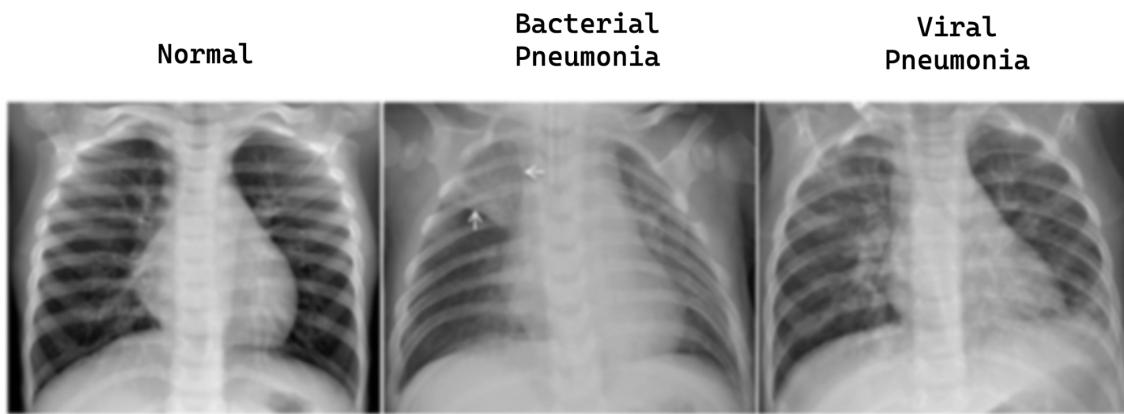
If you're new to computer vision, this project would be the best project that will give you an introduction to this field. The dataset that we'll be using for this project is [the MNIST dataset](#). This data consists of thousands of handwritten digit images from 0 to 9.

Our task is to build a simple deep learning model to predict the digit in a given image of a handwritten digit. This project is ideal for beginners because you'll learn the fundamental image preprocessing steps that you should do before feeding the data to your model. In this project, you'll build a deep learning model consisting of convolutional layers, which is a very common layer to use if we have image data.

If you need step-by-step guidance to do this project, you can follow [this tutorial from Kaggle](#).

Machine Learning Project Idea #11: Detecting Pneumonia Based on Chest X-Ray Result

After finishing your first computer vision problem above, this will be the best follow-up project as it will help you to solidify your knowledge about image classification problems. The data that you will be using is a collection of [chest X-ray images](#).



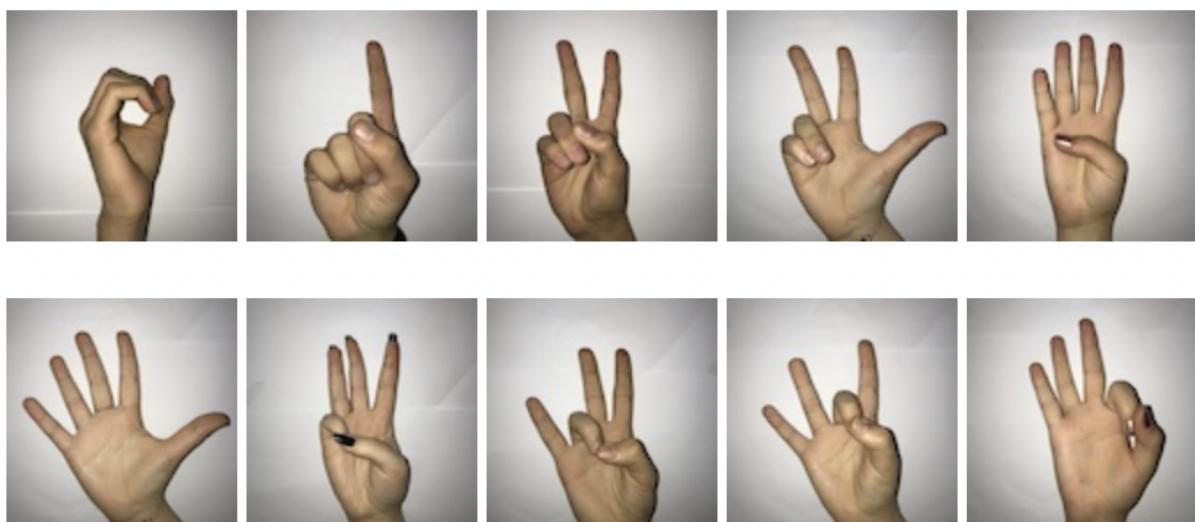
Your task is to build an image classification model to classify whether a patient has pneumonia or not, given the image of an x-ray result. The preprocessing step of using this dataset would be similar to the previous project with handwritten digits.

What's interesting about this project is that the classification task wouldn't be as trivial as the handwritten digits for our machine learning model. This means that you most likely will have to fine-tune the hyperparameters of your deep learning model, such as the model's learning rate, the number of layers, the number of convolutional filters, etc. This will be a good project to showcase your fine-tuning skills.

If you need a step-by-step example of how to do this project, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #12: Detecting Hand Gesture

This project is very much similar to the previous projects in the sense that you're still working on image classification problems. The difference is now you're working with RGB images instead of grayscale images. The dataset that you will be using for this project is [the hand gesture dataset](#). This dataset consists of thousands of images of hand gestures that mimic the shape of a digit from 0 to 10, as you can see below:



Your task in this project is to build an image classification model that is able to take RGB images as inputs and then predict the digit shown in the hand gesture image. If you want to take a look at the step-by-step example of how to conduct this project, check out [this tutorial from Kaggle](#).

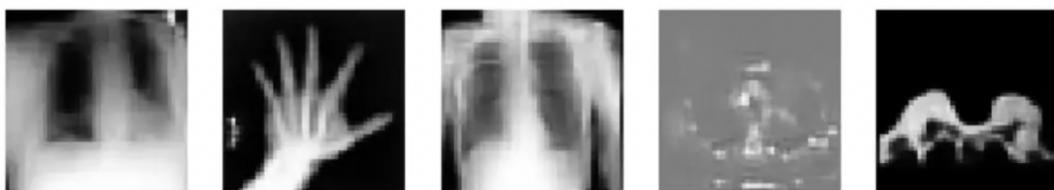
Machine Learning Project Idea #13: Image Denoising with Autoencoder

So far, you've been dealing with mainly image classification projects. Now it's time for you to take your knowledge in computer vision to the next level. The main goal of this project is to clean up the noise in an image, as you can see in the following visualization:

Original noisy images



Reconstructed clean images



The dataset that you'll be using for this project will be the handwritten digit dataset, similar to the dataset that we use in the project 'Classifying Handwritten Digits' above. However, there will be the addition of random noise for each of the images. The noisy images will be the input, and the clean images will be the label to train our image-denoising model.

This project will be helpful to showcase your skill in building an encoder-decoder-based deep learning model, which has different model architectures compared to image classification models.

The encoder portion of the model will be similar to classification models in the sense that we're using convolutional layers to extract features, and then the decoder portion of the model will be the opposite architecture of our encoder to reconstruct our image into a clean one.

If you need step-by-step guidelines on how to conduct this project, check out [this tutorial from Medium](#).

Machine Learning Project Idea #14: Classifying Movie Review Sentiment

In the previous projects about deep learning, we've been dealing mainly with computer vision problems. In this project, we are going to start building a project related to another popular application of deep learning, which is Natural Language Processing (NLP).

The best project to get started in NLP is within text classification domains. For this reason, we're going to use [the IMDB movie review sentiment dataset](#). This would be the best project for you to get started with text classification problems as you can straight away practice the necessary text preprocessing steps before feeding them into your deep learning model.

The goal of this task is to build a deep learning model that can classify the sentiment of a review: whether it is a bad review (negative) or a good review (positive). If you need guidance on how to do this project, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #15: Classifying Tweet Sentiment

This project is similar to the previous project, but instead of predicting the sentiment of a review, we're going to predict the sentiment of a tweet. The exciting part of this project is that you get the chance to collect your own data with Twitter API. Once you collect your data, then we can preprocess the tweets with similar steps as the previous project.

With this project, you have the chance to showcase not only the data preprocessing and modeling part of a machine learning lifecycle but also the data ingestion part. If you need a guideline on how to ingest the data that you need for this project, as well as build a model to classify each tweet's sentiment, refer to [this tutorial on Medium](#).

Machine Learning Project Idea #16: Text Keyword Detection

So far, we've been dealing with text classification problems within the Natural Language Processing use cases. Now in this project, you can do something different that will take your machine learning skills in NLP to the next level.

The main goal of this project is to fetch the keyword in any given text. You can use the same dataset that you've used in previous projects about text classification. There are a lot of open-source keyword extraction methods that are commonly applied in practice, such as TF-IDF, [YAKE](#), [RAKE](#), or [KeyBERT](#).

You can use all of those algorithms to extract the keyword of some texts and play around with the hyperparameters that each algorithm provides. In the end, you can write up some summary about the plus and minus of each algorithm based on your findings.

This project will be beneficial to showcase your analytics skill and how you can gauge some pros and cons by implementing certain algorithms.

Machine Learning Project Idea #17: Topic Modeling of a Text

Topic modeling refers to the task of finding possible topics in a long text. However, different from text classification, topic modeling finds the possible topics in an unsupervised way, meaning that we don't need to provide a label before we feed the text into our machine learning model.

Same to the previous project about keyword detection, there are a lot of topic modeling algorithms that are commonly applied in real-life use cases, such as [Latent Dirichlet Allocation \(LDA\)](#) and [BERTopic](#).

You can use the dataset that you have used for one of the text classification problems above and then try to fine-tune the hyperparameters that each algorithm provides. Finally, you can write a report about the pros and cons of the topic modeling result by using those algorithms.

Advanced Level Project Ideas

ADVANCED LEVEL

PROJECT IDEAS



If you've done some of the projects that are already mentioned above, you should already be familiar with various machine learning algorithms as well as some deep learning models for computer vision and NLP use cases by this stage.

Now in the advanced-level project ideas, we're going to continue to expand our understanding of deep learning. There are several ways we can do this, such as implementing the inner workings of machine learning algorithms and deep learning models from scratch, as well as getting our hands dirty with the application of state-of-the-art deep learning models such as BERT or UNet.

Machine Learning Project Idea #18: ML Algorithms Implementation From Scratch

In the beginner and intermediate-level machine learning projects, we have implemented various machine learning algorithms, such as linear regression, logistic regression, decision tree, SVM, etc., to do classification or regression tasks. However, we implemented those algorithms with the help of established machine learning libraries like Scikit-learn.

To show prospective employees that we have an advanced understanding of the math behind any machine learning algorithms, why don't we implement those algorithms from scratch? Implementing various machine learning algorithms from scratch will not only

impress the prospective employees but will also solidify your understanding of how machine learning algorithms work under the hood, as well as your software engineering skills.

If you need examples and guidance on how to implement various machine learning algorithms from scratch, check out [this GitHub page](#).

Machine Learning Project Idea #19: Backpropagation Implementation of a Simple Deep Learning Model From Scratch

After you have implemented machine learning algorithms from scratch, now you can step your skill up by implementing the algorithm behind a tiny deep learning model from scratch, from the forward propagation method to the backpropagation method.

So far, when we build a deep learning model, we always use established deep learning frameworks such as Tensorflow or Pytorch. However, in this project, we will try to build a tiny deep learning model from scratch.

The main challenge of deep learning implementation is backpropagation, in which the knowledge of algebra and calculus is necessary. Thus, by implementing a backpropagation algorithm from scratch, you will showcase your understanding of the math necessary for deep learning models. Plus, understanding the backpropagation process will make it easier for you to debug the model.

If you need step-by-step guidance on implementing a simple backpropagation process of a deep learning model, check out [this Youtube video from Andrej Karpathy](#).

Machine Learning Project Idea #20: Stock Price Prediction

We all know that the stock market is highly volatile and dynamic. Thus, it's not a secret that predicting a stock price using machine learning has been a challenging task. What we can do in this project is to compare the performance of our machine learning model with a traditional quantitative method such as Moving Average (MA).

The dataset that we can use to do stock price prediction projects normally comes from [Yahoo! Finance](#), and they have an API that we can use to fetch the stock data of a particular company in Python with *yfinance* library.

There are a lot of models that we can use to predict a stock price, such as XGBoost or LSTM. If you need guidance on how to do this project step-by-step, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #21: News Topic Identification

This project can be considered a text classification problem since our goal for this task is to build a deep learning model that is able to classify a news text into one out of six possible categories. To conduct this project, we can use [the BBC news dataset](#).

However, instead of using a normal deep learning model that we have implemented in the previous section, we can use several Transformers-based models such as BERT, RoBerta, Distilbert, etc., and you can compare the performance of each of those Transformers-based models. Creating projects by implementing Transformers-based models is a great way to show that you know how to apply a state-of-the-art model to solve a problem.

HuggingFace provides us with many ready-to-use Transformers-based models, and if you need a step-by-step tutorial on how to use a BERT model for text classification, we recommend you to check [this tutorial from Medium](#).

Machine Learning Project Idea #22: Named Entity Recognition

If the previous project deals with classification problems, in this project, we will do something different, which is Named Entity Recognition (NER) tasks. NER is a task in NLP to extract the entity of a text. Let's say we want to extract a company name from the text '*Bill Gates is the founder of Microsoft*'. With a deep learning model specially designed for NER tasks, we get '*Microsoft*' as a result.

We're going to use the same Transformers-based model from the previous project, but we need to modify their architecture such that it can be used for NER purposes. The data that we can use for this project is [the CoNLL-2003 dataset](#), which is the dataset that is used as a benchmark for NER tasks.

This project will help you to showcase your expertise in solving NLP problems with state-of-the-art models beyond classification tasks. If you want to do this project and need guidance to do so, check out [this tutorial from Medium](#).

Machine Learning Project Idea #23: Text Similarity

Let's say there is a pair of texts, and we want to answer the following questions: how similar are these two texts? Is one text the right translation of another text?

Those questions can be answered by Transformers-based models that we have implemented in previous projects. Thus, the goal of this project is to build a Transformers-based model to predict the similarity of any pair of sentences.

We will get the opportunity to train the model with our own data, which is [the STSB multi-dataset](#). The exciting part of this project is that we finally need to construct our own loss function.

As you might know, when we build a deep learning model for classification purposes, we oftentimes use either binary cross-entropy loss or categorical cross-entropy loss. Meanwhile, in this project, we will have the opportunity to implement a contrastive loss function to push unrelated pairs of texts far apart whilst maintaining a close distance between similar pairs of texts.

If you want to learn how to do this project, check out [the Sentence Transformers library designed specifically for this task](#).

Machine Learning Project Idea #24: Question-Answering

In this project, we will also be using Transformers-based models to build a deep learning model that is able to generate a textual answer given a textual question and the context. Question-answering model's application is very common in real-life use cases, as it often helps with automating the answer to frequently asked questions (FAQ).

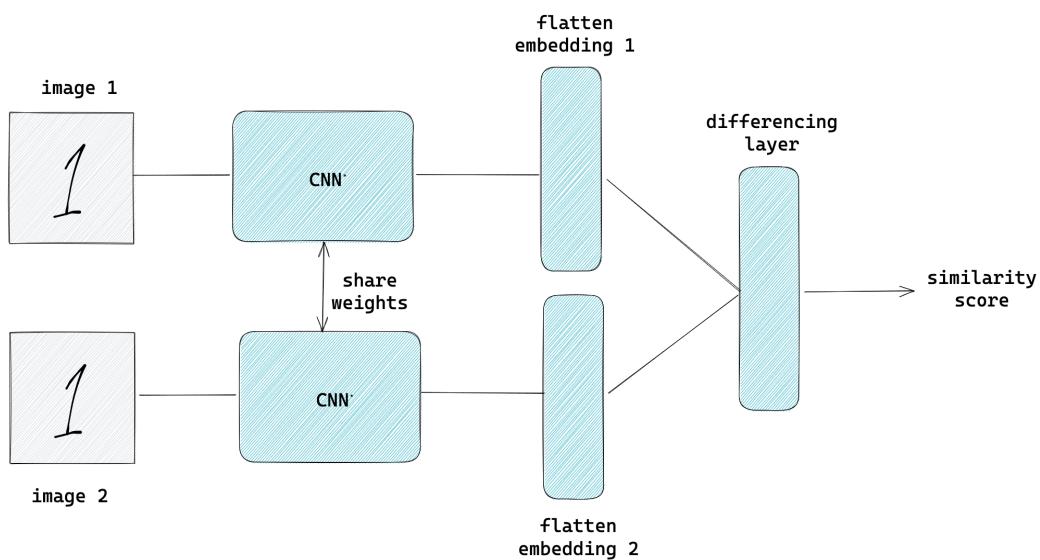
This project can also be done with Transformers-based models from Hugging Face, just like the previous projects. We also have the opportunity to showcase our skills to fine-tune a state-of-the-art model by training it on our own dataset. The dataset that we can use for such a Question-Answering task would be [the Squad dataset](#), which consists of around 100,000 question-answer pairs.

If you need guidance on how to do this task, check out [this guide from HuggingFace](#).

Machine Learning Project Idea #25: Image Similarity

So far, we have dealt mainly with NLP problems. Now it's time for us to do projects related to computer vision problems. This project is similar to the 'Text Similarity' project in the sense that we're dealing with similarity problems. However, instead of assessing the similarity between two texts, we're going to assess the similarity between two images.

For this project, we can build a CNN-based deep learning model with a siamese architecture, as you can see in the following visualization:



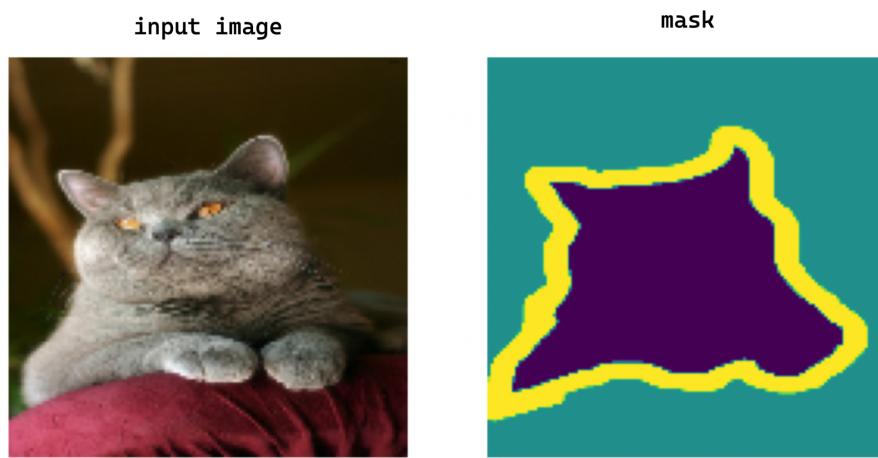
The loss function would be similar to the previous project with text similarity, in which the purpose of the loss function would be to push the distance of dissimilar images far apart and keep similar images close. For the dataset itself, we can use any dataset that we have used for classification purposes in the previous tasks. However, we need to convert the data such that we have pairs of images instead of individual images and then label them accordingly (i.e., 0 if a pair of images doesn't belong to the same class, and 1 otherwise)

If you need guidance on how to build your first Siamese network for image similarity purposes as well as how to prepare your data such that you can use it for image similarity purposes, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #26: Pet Breeds Recognition

One of the most common applications of computer vision is image segmentation, which is very useful for automated tasks such as face recognition, medical image analysis, video surveillance, autonomous vehicles, etc. The main goal of an image segmentation task is to segment objects found in an image and label each of the segments accordingly.

In reality, data labeling of image segmentation tasks takes a lot of time. Thus, in this project, we can use an established dataset, which is [the Oxford-IIT Pets dataset](#). This dataset consists of 37 distinct pet breeds, and each breed consists of around 200 images.



As for the deep learning model, we'll be implementing the most famous models for image segmentation purposes, which is the UNET model that consists of encoder blocks and decoder blocks.

If you need guidance on how to do the project step-by-step, check out [this tutorial from Kaggle](#).

Machine Learning Project Idea #27: Language Translation

Machine translation refers to a task in machine learning in which the model can translate one text into another text in another language. Thus, as you might imagine, we're going to use one of the Transformers-based models for this task, such as the Marian model. This model has been pre-trained to translate English texts into French.

It will be very time-consuming to train a language translation model from scratch, hence what we can do in this project is fine-tune the Marian model on a custom dataset, such as [the KDE4 dataset](#).

If you need step-by-step guidance on how to do this project, check out [this great tutorial from HuggingFace](#).

Machine Learning Project Idea #28: Image Generation

Image generation refers to the task in Computer Vision in which a deep learning model can generate images. Many algorithms are commonly used for this kind of task, such as Generative Adversarial Networks (GANs), Variational Autoencoder (VAE), and the most recent one is the diffusion model.

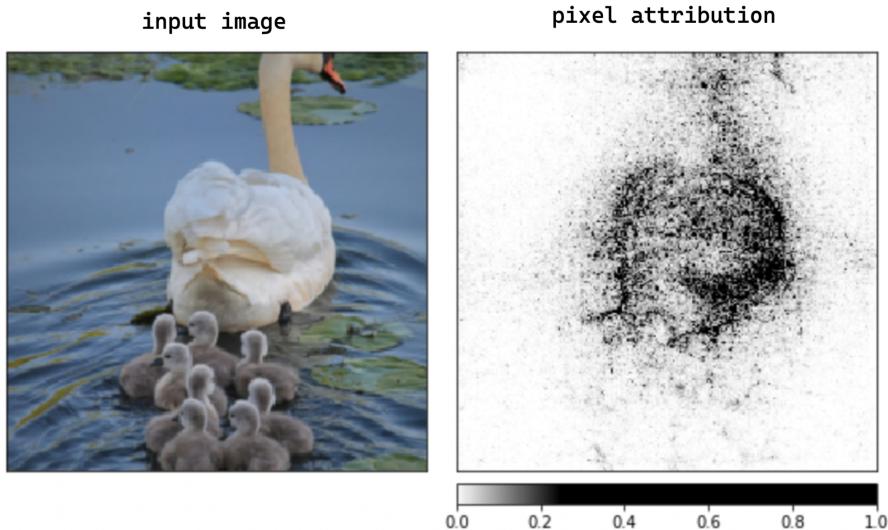
For this project, we can use a wide variety of datasets, including the one that we have used for computer vision tasks in previous projects. However, it's always better to start with a small and simple dataset, such as the handwritten digits dataset, CIFAR10 dataset, etc. Next, we can train a simple GAN model, VAE model, and diffuser model and then compare the image generation results between the three of them.

If you want to know how to train your own GAN model, we recommend you to check out [this tutorial from PyTorch](#) or if you need guidance on how to train your own VAE model, check out [this tutorial](#).

Machine Learning Project Idea #29: Interpreting Prediction of Deep Learning Model with Integrated Gradients

Now that you know how to implement various deep learning models for tasks in computer vision and NLP, let's take another step to showcase your skill in deep learning. We all know that deep learning models consist of a deep stack of layers and thus, they are well known as a black box model - it's not easy to interpret the meaning behind their predictions.

However, recently there are methods that can be used to interpret the prediction of a deep learning model, and the Integrated Gradients method is the most popular one. Currently, there is an established library that allows you to implement those methods, such as [Captum](#) if you build your deep learning model with PyTorch.



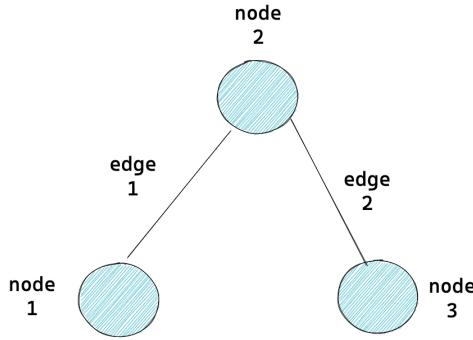
This is a perfect project to show your curiosity about why our deep learning model behaves the way it does after we trained it. Plus, interpreting the prediction of our model is a great way to assess the robustness and reliability of our model before we use it in a production setting.

If you need a step-by-step guide on how to do this project, we recommend you to check out [the official website of Captum](#), as they have perfect tutorials for several use cases.

Machine Learning Project Idea #30: Research Paper Classification

All of the deep learning models' architecture that we have implemented on our projects consist of layers commonly used for computer vision, NLP, or time series forecasting. In this project, we will be implementing another type of deep learning model, which is a graph-based deep learning model called graph neural networks.

As the name suggests, graph neural networks (GNN) expect our input data to be in the form of graphs. Thus, the main challenge of implementing GNN is the way we transform our data such that it can be represented in the form of a graph. In computer science, a graph is a data structure that has two main components: nodes and edges, as you can see in the following visualization:



As an introduction to GNN, our goal in this project is to build a graph-based neural network that is able to classify the category of a scientific publication journal given its citations to other journals with the help of PyTorch Geometric. The dataset that we use for this project is [the Cora dataset](#) which consists of 2708 scientific journals that belong to 7 different categories, with 5429 citation links among them. Each journal represents a node, and each citation link acts as an edge.

If you need guidance on how to do the project, check out [the following tutorial from PyTorch Geometric](#).

Machine Learning Project Idea #31: Text Summarization

Some of us don't really like to read a long-form article and prefer to read the TLDR section of it, so why don't we create our own project to enable us to do exactly that? It turns out that machine learning can be used to extract the summary of a text.

There are many challenges or methods that we can do in order to build a text summarizer model. One possible workflow is that we can split the text into sentences, build a similarity matrix, generate rank based on the similarity matrix, and pick the top N sentences. This, in turn, will make it a project that you can do to showcase your skills within the NLP area.

We can use the *networkX* library to do this project, and if you need a step-by-step example of how to do this project, we recommend you to check out [this tutorial on Medium](#).

Conclusion

In this article, we have seen some project ideas that can showcase your skills in machine learning, starting from the beginner level to the advanced level. Having a

career in data science and machine learning requires us to learn about new stuff almost every day, as the advancement rate of this subject of study is really fast.

Continuously doing personal projects by applying the latest state-of-the-art models is one of the best ways to keep us up to date with the latest advancement within the area of machine learning. If you're on the hunt for a machine learning role, personal projects will give you that unique factor that can set you apart from hundreds of other candidates.