**Credit Card Fraud Detection Project**

Finding anomalies in credit cards and detecting fake credit card transactions was challenging before the advent of machine learning. This project will approve and differentiate swindling credit card transactions from legitimate ones. This project will make you discover and learn how to perform the classification of data. Also, you have to be thorough with the concepts like decision trees, artificial neural networks (ANN), logistic regression, and gradient-boosting classifiers. You can implement the [credit card fraud detection project](https://www.projectpro.io/project-use-case/anomaly-detection-using-deep-learning-and-autoencoders?utm_source=MLStuBlg&utm_medium=Prolink) using libraries like NumPy, Pandas, Matplotlib, Seaborn, XGBClassifier, and frameworks like Scikit-Learn. The[credit card dataset](https://datahub.io/machine-learning/creditcard#readme) and[credit-card fraud detection dataset](https://www.kaggle.com/mlg-ulb/creditcardfraud) are preferable to train your ML model for such a project. These datasets have credit card details along with dummy data of fraudulent and non-fraudulent transactions.

#### ****Fake News Detection Project****

It is another innovative machine learning project for final year students. As you all know, fake news is speeding like a bushfire. Right from connecting people to reading the daily news, everything is available on social media. Hence, it has become more skeptical these days to detect fake news. Many popular social media platforms like Facebook and Twitter already have fake news detection algorithms running behind the scenes on posts and feeds. Implementing this kind of ML project requires good know-how of [various NLP techniques](https://www.projectpro.io/article/10-nlp-techniques-every-data-scientist-should-know/415) and classification techniques (PassiveAggressiveClassifier or Naive Bayes classifier) to detect fake news. PassiveAggressiveClassifier is an online learning algorithm that remains passive while discovering correct classification outcomes. You can prefer supervised learning models to develop such projects.

Libraries like NumPy, Pandas, Itertools, and spaCy (for NLP tasks) will become handy for such a project. Apart from that, frameworks like Scikit-Learn and Streamlit will help a lot. Scikit-Learn has different machine learning approaches and statistical modeling for clustering, classification, and regression. Streamlit helps in building web applications more efficiently and quickly and includes the deployment of ML models. Also, the[Great Fake News dataset](https://github.com/derevirn/gfn-detector/tree/main/data) and[ISOT Fake News Dataset](https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php) are the best for training this ML model.

#### 1. Recommender System Projects

Have you ever seen movies or web series on online streaming platforms? Once you watch one or two of them, you will notice that apps like Netflix and Amazon Prime recommend new web series and movies. It is because these apps render [machine learning models](https://www.projectpro.io/article/common-machine-learning-algorithms-for-beginners/202) that try to understand the customer's taste. Modern e-commerce sites like Flipkart, Amazon, Alibaba, etc., also have the same feature. Recommendation engines are popular in media, entertainment, and shopping. All modern apps come with a recommendation engine that suggests users for more engagement.

You can use libraries like recommenderlab for testing and [developing the recommendation models for your ML project](https://www.projectpro.io/article/recommender-systems-python-methods-and-algorithms/413). Apart from that, libraries like ggplot, reshape2, data.table will complement your machine learning project. Datasets like Google Local, Amazon product reviews, MovieLens, Goodreads, NES, Librarything are preferable for creating recommendation engines using machine learning models. They have a well-researched collection of data such as ratings, reviews, timestamps, price, category information, customer likes, and dislikes.

#### 2. Sales Forecasting Project

      Big B2C marts and retailers want to predict the sales demand for each product present in their inventory. Sales forecasting helps business owners get a clear idea of what products are in demand. Accurate sales forecasting will reduce wastage to a significant level and determine the incremental impact on future budgets. Retailers like Walmart, IKEA, Big Basket, Big Bazaar leverage sales forecasting for sale predictions of product requirements.

To build such ML projects, you must know different approaches to cleaning raw data. Also, must have a thorough understanding of [regression analysis](https://www.projectpro.io/article/types-of-regression-analysis-in-machine-learning/410) especially, simple linear regression. You have to use libraries like Dora, Scrubadub, Pandas, NumPy, etc., for developing these kinds of projects. Dummy datasets like univariate time-series datasets, shampoo sales datasets, etc., can help you model such [machine learning projects](https://www.projectpro.io/article/why-you-should-learn-machine-learning/362).

#### 3. Stock Price Prediction Project

Creating a stock price prediction system using machine learning libraries is an excellent idea to test your [hands-on skills in machine learning](https://www.projectpro.io/article/top-10-machine-learning-projects-for-beginners-in-2021/397). Students who are inclined to work in [finance](https://www.projectpro.io/article/how-data-science-in-finance-has-increased-the-industrys-profitability/169) or fintech sectors must have this on their resume. Nowadays, many organizations and firms lookout for systems that can monitor, analyze and predict the performance and stock price. There is a broad spectrum of data available on finance and the stock market. For this reason, the final year students find it a hotbed of opportunities. Before you start working on this project, you must have proficiency in the following areas:

a.      Statistical modeling: Here, the student has to prepare a real-world model of the mathematical representation and the statistics uncertainties within the analyzing and prediction process.

b.      Regression analysis: This technique talks about the predictive methods that your system will execute while interacting between dependent variables (target data) and independent variables (predictor data).

c.      Predictive Analysis: This analysis will utilize data mining, web scraping, and data exploration techniques for better prediction and accurate analysis.

For analyzing statistical data and handling data clusters in the[stock price prediction project](https://www.projectpro.io/project-use-case/time-series-forecasting-1?utm_source=MLStuBlg&utm_medium=Prolink), you must use libraries like Sklearn, SciPy, Pandas, and SciPy. If you want to visualize the data, you can use Seaborn or Matplotlib. For monitoring and visualizing analyzed stock price and stock market data, you can use Tableau. For training the machine learning model, you can use the[NSE-TATA-GLOBAL dataset](https://github.com/mwitiderrick/stockprice/blob/master/NSE-TATAGLOBAL.csv). It contains all the attributes you need to build your stock price prediction system.

#### 5. Patient's Sickness Prediction System

Machine learning has been proven effective in the field of healthcare also. Traditional healthcare systems became increasingly challenging to cater to the needs of millions of patients. But, with the advent of ML, the paradigm shifted towards value-based treatment. Every modern healthcare equipment and gadget comes with internal apps that can store patient's data. You can leverage these data to create a system that can predict the patient's ailment and forecast the admission. KenSci is an AI-based solution that can analyze clinical data and predict sickness along with more intelligent resource allocation.

            You can use open-source medical datasets like CHDS (Child Health and Development Studies), HCUP, Medicare to test your machine learning algorithm. You can incorporate such projects in healthcare wearables, telemedicine, remote monitoring, etc. To develop such algorithms, you need to have a thorough understanding of the following:

a.      Classification & Clustering model: While classification determines the categorization of data, clustering in ML looks for distinctive patterns in the data when the data available does not have a definite outcome.

b.      Regression analysis: Its principal purpose is to find value. This technique talks about the predictive methods that your system will administer while interacting between dependent variables (target data) and independent variables (predictor data).

 Use libraries like NumPy, Pandas, Matplotlib, Theano, etc., and frameworks like Keras and Hugging face to implement this ML project.

#### 7. Email Spam-Filtering System

Mining text is one of the popular computation techniques widely applied in applications like text summarization, topic classification, machine translation, sentiment analysis, etc. Modern cybersecurity systems are utilizing machine learning methods a lot. Spam email detecting systems are one of them. Spam filtering also leverages text mining and document classification to segregate legitimate mails and spam emails. All modernized email services come with this segregation system that runs machine learning algorithms behind. Such a project comes under the text [classification problems](https://www.projectpro.io/article/7-types-of-classification-algorithms-in-machine-learning/435). Building this kind of a ML project involves the following important steps -

a.      Text Processing

b.      Text Sequencing

c.      Model Selection

d.      Implementation

Use libraries like Sklearn, NumPy, Counter, Scrubadub, Beautifier, Seaborn, and machine learning frameworks like TensorFlow and Keras. For training such machine learning models, a[Spambase dataset](http://archive.ics.uci.edu/ml/datasets/Spambase/) can help you. Spambase dataset is an open-source UCI machine learning repository comprising around 5569 emails, of which nearly 745 are spam emails.

#### ****. Digit Classification Project using MNIST Dataset****

The digit classification project is a remarkable machine learning project that employs [neural network and machine learning concepts](https://www.projectpro.io/article/5-different-types-of-neural-networks/431). From the outset of machine learning, it was challenging to work with unstructured data (image dataset) and transform it into structured data (texts). In this project, you will use [Convolutional Neural Networks](https://www.projectpro.io/article/rnn-vs-cnn-the-difference/491) (CNNs) to train machine learning algorithms. The MNIST dataset will make the training of ML models seamless so that your system can recognize handwritten digits easily.

  This type of project comes under the domain of computer vision. To build this project, you have to use libraries like NumPy, PIL, Pillow, Scikit-image, Tkinter, etc. Tkinter will help in developing a GUI interface for your application. All the neural network algorithms will get managed by Keras. You have to use frameworks like TensorFlow and Keras. You can train your model with the[MNIST dataset](http://yann.lecun.com/exdb/mnist/) having sixty thousand images containing handwritten digits from zero to nine for training and around ten thousand images for testing purposes.

#### 13. Music Genre Classification System

It is an advanced machine learning project that leverages deep learning concepts to classify music files based on different genres. Many music companies like Spotify, Gaana use such algorithms in their app to understand the listening preferences of the user. You can also select a specific genre for your customized playlist through this project. It uses [deep learning algorithms](https://www.projectpro.io/article/deep-learning-algorithms/443)to classify the list of songs to the smartphone user. To develop this project, you need to convert the audio signals into compatible formats. You need to have an understanding of two important concepts -  Spectrogram generation and Wavelet generation.

            This project also requires K-Nearest Neighbor Algorithm (KNN), Convolution Neural Network (CNN), and [Support Vector Machines](https://www.projectpro.io/data-science-in-r-programming-tutorial/support-vector-machine-tutorial) (SVM) to develop the CNN model. Then we have to pass the spectrogram & wavelet data in that CNN model for multimodal training. This project will demand an audio dataset. So, you can use the[GITZAN dataset](http://opihi.cs.uvic.ca/sound/genres.tar.gz) that comprises 1000 music files. Also, this dataset has ten different varieties of genres (blues, hip-hop, metal, classical, disco, country, pop, rock, jazz, and reggae) included with uniform distribution. Each of these files has a length of thirty seconds.

#### 1. Cartoonify Image with Machine Learning

**Project Idea:** Transform images into its cartoon. Yes, the objective of this machine learning project is to CARTOONIFY the images.

Thus, you will build a python application that will transform an image into its cartoon using machine learning libraries.

#### 2. Iris Flowers Classification Project

**Project idea –** The iris flowers have different species and you can distinguish them based on the length of petals and sepals.

This is a basic project for machine learning beginners to predict the species of a new iris flower.

**Dataset:** [Iris Flowers Classification Dataset](https://archive.ics.uci.edu/ml/datasets/Iris)

#### 3. Emojify – Create your own emoji with Python

**Project idea –**The objective of this machine learning project is to classify human facial expressions and map them to emojis.

You will build a convolution neural network to recognize facial emotions.

Then you will map those emotions with the corresponding emojis or avatars

<https://data-flair.training/blogs/create-emoji-with-deep-learning/>

#### 4. Loan Prediction using Machine Learning

**Project idea –** The idea behind this ML project is to build a model that will classify how much loan the user can take.

It is based on the user’s marital status, education, number of dependents, and employments. You can build a linear model for this project.

**Dataset:**[Loan Prediction Dataset](https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-dataset)

<https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset>

#### 5. Housing Prices Prediction Project

**Project idea –** The dataset has house prices of the Boston residual areas.

The expense of the house varies according to various factors like crime rate, number of rooms, etc.

It is a good ML project for beginners to predict prices on the basis of new data.

**Dataset:**[Housing Price Prediction Dataset](https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html)