

Simple Project Ideas to kickstart your ML journey.

10 Data Science Project Ideas For 2025

Covers Various Domains and Industries.

Disclaimer: Most of these are beginner-intermediate projects.



Nirmal Budhathoki



01

Security: Phishing Email Detection

Description: Build a model that detects phishing emails by analyzing email content and headers. Use NLP techniques to classify emails as phishing or legitimate.

Data Source:

<https://www.kaggle.com/datasets/subhajournal/phishingemails>



Nirmal Budhathoki

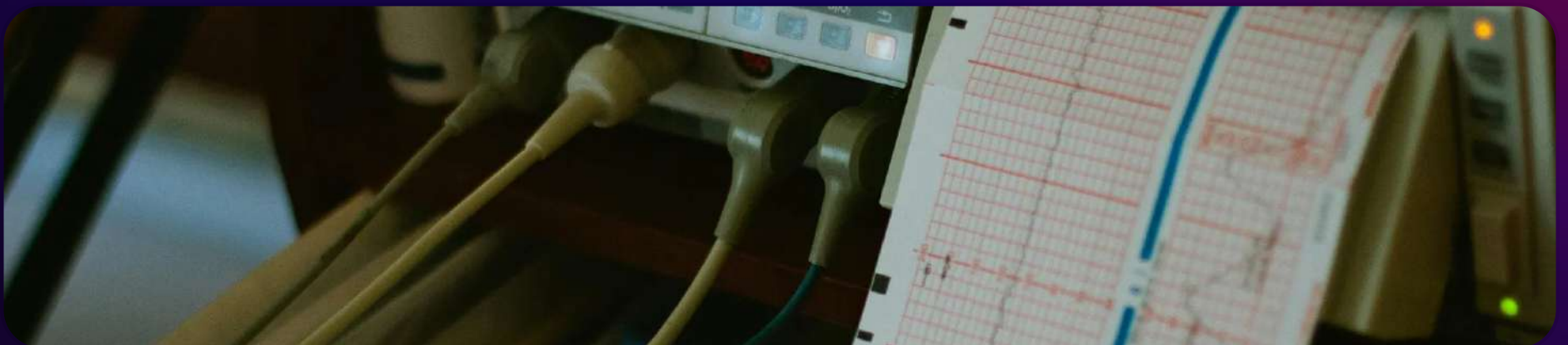
02

Healthcare: Predicting Patients Readmissions

Description: Use patient demographics, diagnoses, treatments, and discharge summaries to predict the likelihood of readmission within 30 days.

Data Source:

<https://archive.ics.uci.edu/dataset/296/diabetics+130-us+hospitals+for+years+1999-2008>



Nirmal Budhathoki

03

Telecom: Customer Churn Prediction

Description: Predict which customers are likely to leave a telecom service provider using features like call durations, customer complaints, billing data, and service usage.

Data Source:

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>



Nirmal Budhathoki

04

RideShare: Demand Forecasting for Rides

Description: Predict ride demand for specific areas based on historical data, weather, and traffic conditions. This can help ride-share companies allocate drivers effectively.

Data Source:

<https://www.kaggle.com/datasets/microize/newyork-yellow-taxi-trip-data-2020-2019/data>



Nirmal Budhathoki

05

e-Commerce: Dynamic Pricing

Description: Develop a pricing algorithm to recommend optimal product prices based on product demand, and customer reviews.

Data Source:

<https://amazon-reviews-2023.github.io/>



Nirmal Budhathoki

06

Auto: Predictive Maintenance

Description: Predict the need for vehicle maintenance using engine data.

Data Source:

<https://www.kaggle.com/datasets/parvmodi/automotive-vehicles-engine-health-dataset/data>



Nirmal Budhathoki

07

Social Networking: Fake News Detection

Description: Build a model that identifies fake news articles shared on social networks by analyzing text and metadata.

Data Source:

<https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification>



Nirmal Budhathoki

08

NLP: Smart Document Classifier

Description: Build a model that classifies legal documents into different categories. MultiClass Document Classifier.

Data Source:

<https://www.kaggle.com/datasets/amohankumar/legal-text-classification-dataset/data>



Nirmal Budhathoki

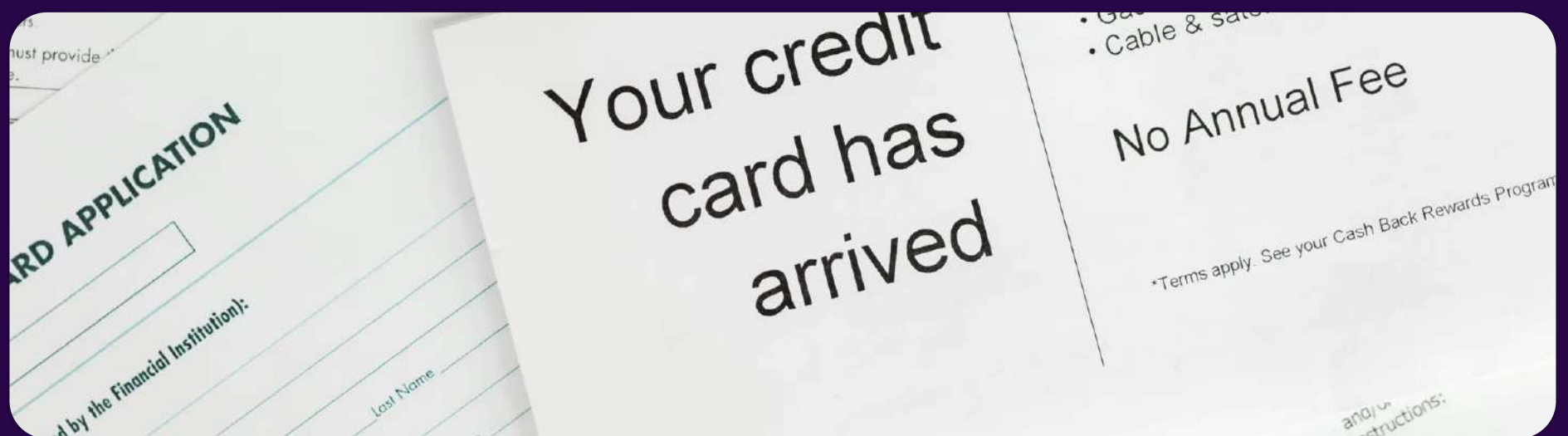
09

Finance: Credit Card Fraud Detection

Description: Detect fraudulent transactions in credit card data by building anomaly detection or classification models.

Data Source:

<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>



Nirmal Budhathoki

10

Environment: Air Quality Prediction

Description: Predict air quality index (AQI) using historical weather data and pollutant levels.

Data Source:

<https://www.kaggle.com/datasets/fedesoriano/air-quality-data-set/data>



Nirmal Budhathoki

Learn More?



Data Sources



Nirmal Budhathoki

