CodSoft: Machine Learning Projects.

This repository contains three machine learning projects completed as part of the CodSoft Internship Program. Each project demonstrates the application of various machine learning techniques to solve real-world problems.

Projects Included:

1. Spam SMS Detection Project

         Objective: To build an AI model that classifies SMS messages as either spam or legitimate (ham).

         Overview: This project leverages natural language processing (NLP) techniques and machine learning classifiers to identify spam messages with high accuracy. The workflow includes text preprocessing, feature transformation, and model evaluation.

         Dataset Name: SMS Spam Collection Dataset

         Techniques Used:

o   NLP: TF-IDF (Term Frequency-Inverse Document Frequency) or Word Embeddings

o   Classifiers: Naive Bayes, Logistic Regression, Support Vector Machines (SVM)

o   Metrics: Accuracy, Precision, Recall, and F1-score

         Result: The model demonstrates high accuracy and effectiveness in automating spam detection for text communications..

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy: 98% | Precision: 99% | Recall:92% | F1-Score: 95% |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

      2.Credit Card Fraud Detection: Key Points

         Objective: Build a machine learning model to detect fraudulent credit card transactions and classify them as either fraudulent or legitimate.

         Overview: This project addresses the critical challenge of identifying fraudulent activities in financial systems by leveraging machine learning techniques. It focuses on analyzing transaction data to ensure accuracy in fraud detection.

         Dataset Name: Credit Card Fraud Detection Dataset

         Techniques Used:

o   Data Preprocessing: Handling missing values and class imbalance (e.g., SMOTE).

o   Feature Engineering: Scaling features with StandardScaler.

o   Algorithms: Logistic Regression, Decision Trees, Random Forests

o   Evaluation Metrics: Precision, Recall, F1-score.

         Result: Delivered a robust solution with high precision and recall, effectively classifying transactions and combating financial fraud.

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy: 96.68% | Precision: 96.94% | Recall: 96.33% | F1-Score: 96.64% |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3.Movie Genre Classification Project

         Objective: To classify movies into their respective genres based on plot summaries or other textual information using machine learning techniques.

         Overview: Build a machine-learning model to automate genre classification. Utilize natural language processing (NLP) to process and analyse text data.

         Dataset Name: Genre Classification Dataset from IMDb

         Techniques Used:

o   Text Preprocessing: Tokenization, stop-word removal, and stemming/lemmatization.

o   Feature Extraction: TF-IDF (Term Frequency-Inverse Document Frequency), Word Embeddings (e.g., Word2Vec, GloVe)

o   Machine Learning Models: Naive Bayes, Logistic Regression, Support Vector Machines (SVM)

o   Evaluation Metrics: Accuracy, precision, recall, and F1-score.

         Result: An efficient and scalable solution for predicting movie genres with high performance across evaluation metrics.

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy: 98% | Precision: 99% | Recall:92% | F1-Score: 95% |