Intermediate Data Science Projects

1. Customer Segmentation using Clustering

Project Focus:
Unsupervised learning using K-Means or Hierarchical Clustering.

Project Goal:
Group customers based on purchasing behavior to develop targeted marketing strategies.

Key Skills:
- Clustering techniques (K-Means, Hierarchical)
- Data Preprocessing
- Data Visualization

Tools and Libraries:
- Python (Pandas, Scikit-learn, Matplotlib, Seaborn)

Dataset:
https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python

Completion Date: 12th Oct 2024

2. Credit Card Fraud Detection with Deployment

Project Focus:
Classification to detect fraudulent credit card transactions and deploy the solution.

Project Goal:
Build a fraud detection system and deploy it as a web app.

Key Skills:
- Classification (Logistic Regression, Random Forest)
- Data Imbalance Techniques (SMOTE)
- Web Development and Deployment

Steps:
1. Preprocess the dataset and build a classification model using Logistic Regression or Random Forest.
2. Use SMOTE or other techniques to handle imbalanced data.
3. Deploy the model** using Flask or Django.
4. Host the app using Heroku, AWS, or similar cloud services.

Tools and Libraries:
- Python (Pandas, Scikit-learn, Imbalanced-learn)
- Flask/Django for web development
- Heroku/AWS for deployment

Dataset:
https://www.kaggle.com/mlg-ulb/creditcardfraud

Completion Date: 22nd Oct 2024

3. Sentiment Analysis on Product Reviews

Project Focus:
Natural Language Processing (NLP) for text classification of customer reviews.

Project Goal:
Perform sentiment analysis on product reviews to classify them as positive, negative, or neutral.

Key Skills:
- Text Preprocessing (Tokenization, Lemmatization)
- Feature Extraction (TF-IDF, Word Embeddings)
- Classification Algorithms (Naive Bayes, SVM)

Tools and Libraries:

- Python (Pandas, Scikit-learn, NLTK, SpaCy)
- Word Embeddings (Word2Vec, GloVe)

Dataset:

https://www.kaggle.com/bittlingmayer/amazonreviews

Completion Date: 02nd Nov 2024