Introduction

A Foray into Financial Data Analysis

Contents

Introduction

This group project focuses on exploring a chosen application domain through the lens of data science. The primary objective is to gain hands-on experience with data science techniques while developing collaborative working skills and expanding our programming knowledge. We have explored different types of datasets and possible application domains. Through this research, we have chosen to focus on finance data, specifically exploring areas such as credit card fraud detection and financial time series analysis.

The primary objective is to gain practical experience with data science techniques within the financial domain, while fostering collaborative teamwork and enhancing our programming skills. The exploration of this application domain will guide our choice of dataset for analysis in Assessment 1.

This project is structured as follows:

- We start with research on the types of datasets and application domains. We then justify our choice of application domain (finance) and show the resources used in exploring and finding datasets. We explore the types of financial data that exist. We search for sources of data and assess their suitability for use to explore a classification or regression problem. This can be found in `02-Data`.
- We explore two datasets: one on credit card fraud (see the files `03-01-CreditCardFraud` and `03-02-CreditCardFraudimbalanced`) and another on Apple stock data (see the file `04-AppleStock`). We explored the ways such data is analysed by viewing existing literature and online resources. We then presented our versions of exploratory data analysis to understand the data structure and understand potential challenges in analysing such data.
- We judge the suitability of such datasets: We will consider the statistical methods that are used to analyse the data and check whether they fall within the scope of Assessment 1.

Throughout, we will explore the use of general data science resources, assessing their applicability to financial datasets. This will help us build a repertory of resources that we can then use during the next assessment, and even further beyond. Our team will identify relevant resources, run and modify code examples, and create original visualizations, culminating in a structured report that synthesizes our findings and insights.

References

1. Project brief, accessible on Github. This link was accessed 18:00 on 26/09/2024.