**Project data management plan**

**Project title:** Future prediction of the bitcoin price

**Research question:**

* What are the key factors influencing bitcoin price?
* How does the trading volume affect the bitcoin price?
* How to predict the future price of bitcoin using historical data?

**Project objectives:**

* Collect and preprocess historical bitcoin price.
* Apply time series analysis to model the data.
* Implement LSTM, ARIMA and Bayesian ridge regression models.
* Compare the accuracy of LSTM, ARIMA and Bayesian ridge regression models.

**Summary of project and background:**

* The project aims to predict future price of bitcoin using machine learning techniques. In this project I intended to predict the price of bitcoin a digital currency that has exhibited significant volatility. If I can predict the price it will help for investors to make a better trade. By of LSTM (Long Short-Term Memory), ARIMA (Autoregressive Integrated Moving Average) and Bayesian ridge regression models, I seek to provide insights into the future price trends of bitcoin.
* **LSTM**: This model is best for complex sequence with patterns over a long period of time.
* **ARIMA**: Best for linear time series with well-defined trends and seasonal patterns. A more traditional and interpretable approach.
* **Bayesian Ridge Regression**: Useful for linear relationships with the added benefit of regularization and uncertainty quantification. Offers a simple yet robust predictive framework.

**Reference links:**

* Chen, J. (2023). Analysis of Bitcoin Price Prediction Using Machine Learning. *J. Risk Financial Manag.*, 16(1), 51. <https://doi.org/10.3390/jrfm16010051>
* Hua, Y., 2023. *Bitcoin price prediction using ARIMA and LSTM*. 1st ed. Jacobs School of Engineering, University of California San Diego, US. <https://www.e3s-conferences.org/articles/e3sconf/pdf/2020/78/e3sconf_iseese2020_01050.pdf>
* Shah, D. and Zhang, K., 2014. Bayesian Regression and Bitcoin. [pdf] Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Department of EECS. Available at: <https://devavrat.mit.edu/wp-content/uploads/2017/11/Bayesian-regression-and-Bitcoin.pdf>

**Data management plan:**

**Overview of the dataset:**

The dataset used in this project consists of historical bitcoin prices. It includes various columns such as Date, Open, High, Low, Close and Volume. For this project the “Close” price are primarily used for the prediction.

**Dataset Source:**

* **Kaggle-** [**https://www.kaggle.com/datasets/yasserh/bitcoin-prices-dataset**](https://www.kaggle.com/datasets/yasserh/bitcoin-prices-dataset)

**Summary of data**

The dataset consists over several years of bitcoin price. And the dataset is in the CSV. file format and the dataset contain 2684 rows and 7 columns. The dataset has less than 500 kilobytes.

**Ethical Compliance:**

This project adheres to ethical standards and ensures under GDRP by not using personal or sensitive data, alignment with University of Hertfordshire’s ethical policies, The Dataset consists of publicly available Bitcoin price (only consists of values). The dataset is only used for Research and Education purpose only.

**Document control:**

All the versions of the data and any changes made during the preprocessing and analysis are documented and stored multiple places and version control is managed throughGit hub( <https://github.com/bm23abb/7PAM2002-0509-2023---Data-Science-Project.git> ), My drive(<https://drive.google.com/drive/folders/1-mpZHclvAZLmHZoutiq7-gPicCkZgw1c?usp=sharing>), hard drive(File name: Data-Science-Project), Tera box(<https://terabox.com/s/12ZeS6DbOp460-gSlRxME5Q>).

**Metadata:**

README.md file in the GitHub repository, providing a project overview, data source and descriptions, codes and contact information are entail in the inclusion.

**Security and storage:**

Ensuring data security and made a proper storage for the integrity and confidentiality of the project.I am planning to cover different platforms to store project data (GitHub ,My drive ,Tera box, Hard drive) and the link are provided in the document control.

**Project Timeline: Gantt Chart**

A graph with colorful lines

Description automatically generated with medium confidence

**Data Sample:**

A table of numbers and letters

Description automatically generated