

GROUP 7

Members:

- Chioma Susan Nwade
- Maha Tahir
- Rui Fang
- Nadiia Sharova
- Yue Liu

Question 1 [20 points]

What is your project question(s) and the problem it tackles?

- You will be marked on how realistic the problem is given the project timeline, as well as the fit with the subjects covered throughout the Data Specialisation and your personal, common interests in the topic.

Project problem

The problem addressed by this project is the lack of accurate and reliable cryptocurrency price forecasting models. The cryptocurrency market is highly volatile and influenced by various factors, including market sentiment, technological advancements, regulatory changes, and macroeconomic events. As a result, investors, financial institutions, and fund managers face challenges in making well-informed decisions to optimize their investment strategies and mitigate risks.

Project question

"How can an accurate and reliable cryptocurrency price forecasting model be developed to assist investors, financial institutions, and fund managers in optimizing their investment strategies, mitigating risk, and making well-informed decisions in the cryptocurrency market?"

Problem Objective

To develop an accurate and reliable cryptocurrency price forecasting model that aids investors, financial institutions, and fund managers in optimizing their investment strategies, mitigating risk, and making well-informed decisions in the cryptocurrency market.

Potential Benefits

Optimizing Investment Strategies: The model can assist investors in optimizing their cryptocurrency investment strategies. By providing precise forecasts of cryptocurrency prices, investors can make informed choices about when to buy, sell, or hold different cryptocurrencies. This optimization can lead to enhanced returns and reduced potential losses, ultimately contributing to improved portfolio performance and overall financial gains.

Risk Mitigation and Decision Support:

The model can play a crucial role for financial institutions and fund managers by helping them mitigate risk and facilitating decision-making processes. With a robust cryptocurrency price forecasting model, these professionals can gain deeper insights into the potential risks linked to their cryptocurrency investments. This predictive tool can serve as a valuable asset for decision-making, empowering them to make timely adjustments or reallocations to prevent significant losses and maintain a strong financial stance.

By achieving the objective, the project aims to empower stakeholders with a valuable tool that contributes to more successful and effective decision-making in the complex and dynamic cryptocurrency market.

Question 2 [20 points]

Explain your target audience. Who could be interested in reading your final report and for whom will your project be useful? Assess the level of expertise in relation to data science of your audience (for example, how technical should your report/presentation be?).

- You will be marked on correctly identifying your audience and the technical level of your presentation and report.

1. Investors:

This group consists of individuals and entities looking to invest in the cryptocurrency market. They could range from beginners with limited understanding of data science to more experienced investors who are familiar with financial models and forecasting techniques but might not be experts in data science. For this audience, our report and presentation focus on delivering insights in a clear and accessible manner, explaining the methodology used in the forecasting model without delving into overly technical details.

2. Financial Institutions:

This category encompasses banks, asset management firms, and hedge funds that are interested in including cryptocurrencies in their investment portfolios. Professionals in this group may have a higher level of financial expertise but might not be deeply involved in data science. So our report highlights the benefits of using the model in their investment strategies, emphasising risk management and decision support. Provide enough technical information to showcase the model's reliability without overwhelming them with intricate technicalities.

3. Fund Managers:

Fund managers oversee investment portfolios and are responsible for making strategic decisions. They might have a moderate level of technical understanding, especially if they have experience in quantitative finance. We would tailor our report to include more technical details about the model's methodology, data sources, and performance metrics. With a highlight on how the model's forecasts can aid their decision-making process and contribute to better portfolio performance.

4. Data Scientists and Researchers:

While not the primary audience, fellow data scientists and researchers might also find our work interesting for its methodology and potential advancements in the field of cryptocurrency price forecasting. For this subset, we can delve deeper into the technical aspects, explaining the algorithms, data preprocessing techniques, and model evaluation metrics in detail.

Considering the varying levels of expertise, our final report and presentation strike a balance between accessibility and technical depth. We will start with a high-level overview of the problem, potential benefits, and the model's capabilities. Then, we gradually introduce the methodology, explaining concepts in a straightforward manner, along with visualisations and examples to make the explanations more intuitive. For the technical audience, we will provide additional details on the model architecture, data preprocessing, feature engineering, and evaluation techniques.

In summary, our project's audience includes investors, financial institutions, fund managers, and potentially data scientists/researchers. The technical level of our presentation and report would be adjusted based on the audience's familiarity with data science concepts, ranging from accessible explanations for less technical individuals to more in-depth technical insights for those with greater expertise.

Question 3 [30 points]

What data sources will you need to answer your project question(s)? Describe any potential issues you can have with the datasets and how will you overcome this:

- For example, will the data you find only cover particular geographical areas? Will you need to combine multiple datasets to overcome this?

The dataset we are using was extracted from <https://coinmarketcap.com/> which is the world's most-referenced price-tracking website for cryptoassets in the rapidly growing cryptocurrency space.

Loading API data into a Jupyter Notebook involves several steps and the following were the issues faced during the ETL process.

API Access: CoinMarketCap required us to create an account and obtain an API key for accessing their data. Managing and protecting API keys is important for security.

Authentication Errors: We had to make sure that the API keys were correct to avoid authentication errors when making API requests.

API Limitations: CoinMarketCap had a usage limits on their API, such as a maximum number of requests per minute or per day. The maximum number of unique cryptocurrencies was limited to 5000. This is the number of observations in our dataset.

Data Format: The API returned the data in json format so we had to convert it into a pandas data frame in order for us to carry out data cleaning.

Data Cleaning and Transformation: We had to drop some redundant/repeated columns, convert datatypes, check for missing values, duplicates, separate texts from numbers and check for outliers. This was all done to make the data to be of good quality for it to be used effectively in our analysis.

Question 4 [30 points]

Describe the team approach to the project work:

- How are you planning to distribute the workload and how are you planning to work on your project;
- What are your team's strengths and weaknesses;
- Include an expected timeline of the project
- How are you managing your code;

We have dividing out project into sections for the next 4 weeks which we feel are important. We have weekly meetings on Fridays and those meetings are recorded and uploaded onto our chat in case anyone needs to reference back.

Within these meetings we discuss the topic, share our understands and update the team about our progress. Team members are also uploading material to help the rest of the team to get onto track.

Sections with timeline being:

- Week 1: (31/07/2023)
 - Define business objectives (Team)
 - Extract data (Chioma)
 - Clean and explore data (Renee)
- Week 2: (01/08/2023)
 - Visualise and understand data (Maha)
 - Analysis (Team)
- Week 3: (14/08/2023)
 - Feature engineering (Yue)
 - Machine Learning Engineering – Model Building (Chioma)
- Week 4: (21/08/2023)
 - Review and construct a PowerPoint (Nadiia)
 - Submission (Team)
- Week 5
 - Prepare notes for presentation
 - Decide who is presenting what from the slides
 - Rehearse as much as we can

This list shows our expected timeline for the project

The reason we have created this list is so that we can ensure we are hitting out weekly targets. This will also allow everyone to see their appointed tasks and in case anyone needs support, we will all chip in as a team to aid.

A Trello account has also been created where we can send our work for the team to review.

Strengths & Weaknesses

Chioma:

+ MSc in Business analytics (distinction): This has equipped her with advanced data handling, ML, analytics, and business acumen. She can communicate insights effectively, and bridge technical and business perspectives.

+Critical Thinking: Your analytical mindset enables you to critically evaluate model performance, understand limitations, and propose adjustments for improvement.

- Tends to become self-critical when I put in a lot of effort into tasks, and it doesn't go according to plan, but I have worked on redirecting my focus towards the lessons learned from these experiences over the years.

- She sometimes allocates considerable time to learning new things during a project, which can reduce her overall productivity time, but she is learning to always strike a balance.

Nadiia:

+ Studied at the Faculty of Architecture and Civil Engineering, specialising in the technology of building structures, products and materials

+ I see the problem from the outside, I can work in a team, I am multitasking, I can distribute tasks, and I also know how to complete tasks, I learn quickly and always achieve my goal

- When working in a rush, important calculations are missed, so I need to inform team to always double check

- Sometimes can take more work on to take load off others, however end up being stressed

Renee:

+ Attention to Detail and Precision: I take pride in being meticulous and thorough in my work. Whether it's a project or a task, I always strive to give my best and pay close attention to even the smallest details. I have a knack for spotting nuances and have a strong insight into the finer points of things, which helps ensure quality and accuracy

+ Strong Logical and Analytical Skills: I have a natural inclination towards logical analysis and reasoning. I excel at breaking down complex problems, analyzing various aspects, and drawing meaningful conclusions. I'm good at summarizing information and honing in on the key aspects of an issue, which helps me tackle challenges efficiently and make informed decisions.

- Perfectionism and Procrastination: I tend to be a bit of a perfectionist, which sometimes makes me delay starting tasks until I feel everything is just right. I want things to be perfect, but I realise this can lead to procrastination. I'm working on finding a balance between aiming for excellence and getting started in a timely manner.

- Slower Initial Learning Curve: I've noticed that I often take a bit more time to really grasp the basics when I start something new. It can be a bit frustrating initially, but once I get a handle on the core concepts, I tend to pick up speed and perform better. I'm trying to find ways to speed up that initial learning curve without sacrificing my understanding of the fundamentals.

Maha:

+ A Biomedical Scientist

+ Putting together material and delivering content, teaching it back

- Working with opinionated people makes me feel like I disappear in the shadows as I feel as though my opinion will not matter or be heard so I keep quiet

- Can get carried away with work whilst trying to perfect it, I sometimes don't know when to stop and disassociate myself from my work life

As a whole team:

+ We are all knowledgeable in different specialities

+ We all have extensive working background so we all have transferable skills to help work together, meet deadlines and communicate effectively

- We are all working full time so sometimes not all of us can make it to the meetings

- We may not understand the topic or area of subject, all at the same time

We are managing the code as a team. Each team member will be creating a code for their area of the task. Some of the ways we will be making the code include:

- Code must have easy readability, this can be achieved by writing as few lines as possible, using appropriate name conventions and by avoiding using lengthy codes for a simple function,
- Having headers to make it easier to understand and maintain the code
- Using Exception handing to catch out and avoid errors, issues or mistakes within the code
- Using the correct Machine Learning, API and codes
- Our code needs to be easily accessible, easy to read, easy to reproduce and easy to maintain

So overall in a nutshell, the code must be clean, consistent, functional and useful to its purpose, easy to understand for its end users, efficient and concise, easy to test and be tested for any defects, easy to maintain and finally easy to replicate, reuse and to add on more code if needed.