# Data Preparation and Analysis (CSP 571)

1. **Project Group:**

| Name | Hawk-id | Email-id |
|---|---|---|
| **Sai Yeshwanth (lead)** | A20514173 | Sbolikonda@hawk.iit.edu |
| Ashlesh Khajbage | A20517913 | akhajbage@hawk.iit.edu |
| Chandana poloju | A20527866 | cpoloju@hawk.iit.edu |
| Swastika Pandit | A20503284 | spandit3@hawk.iit.edu |
| Deepthi pinninti | A20513400 | spinninti@hawk.iit.edu |

## Project Topic:

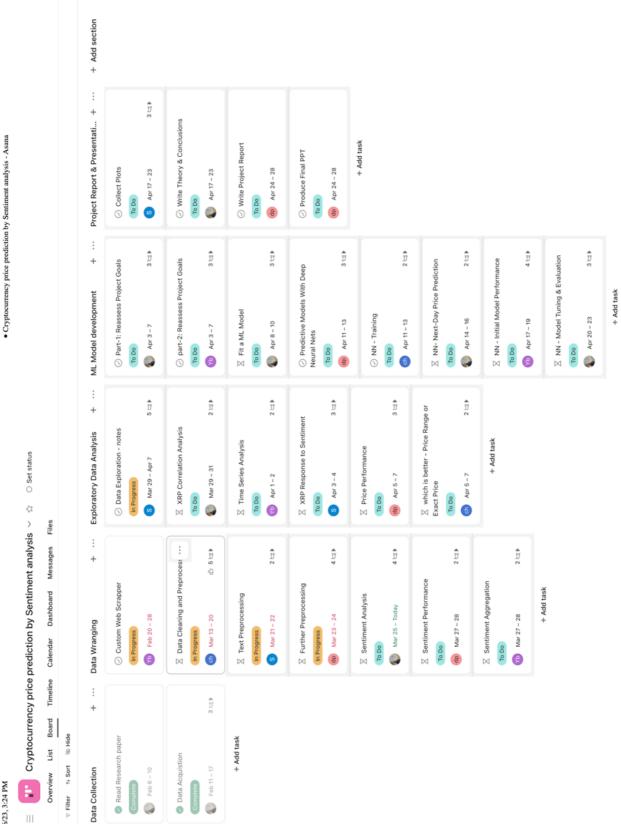## Facilitation of Cryptocurrency price prediction by Sentiment analysis.

Application subject area – **Cryptocurrency**

## What we seek to address

- Does historical data of BTC, LTC, status, or ratings influence this result?
- How keywords in daily news articles impact coin's trading performance.
- Is the Global sentiment library suitable for crypto sentiment analysis? If not, do we need to generate our own library?
- How does the Density Distribution of Deviation of High Price differ from Open Price?
- Can we discover if there is a correlation between investors and R.O.I?
- Do the total traded Quantity and Close Price of the coin hold any sort of relation between them?
- Which analysis model will be best for the analysis of the dataset?
- Can we analyze price movements and trading volumes and give next-day predictions?
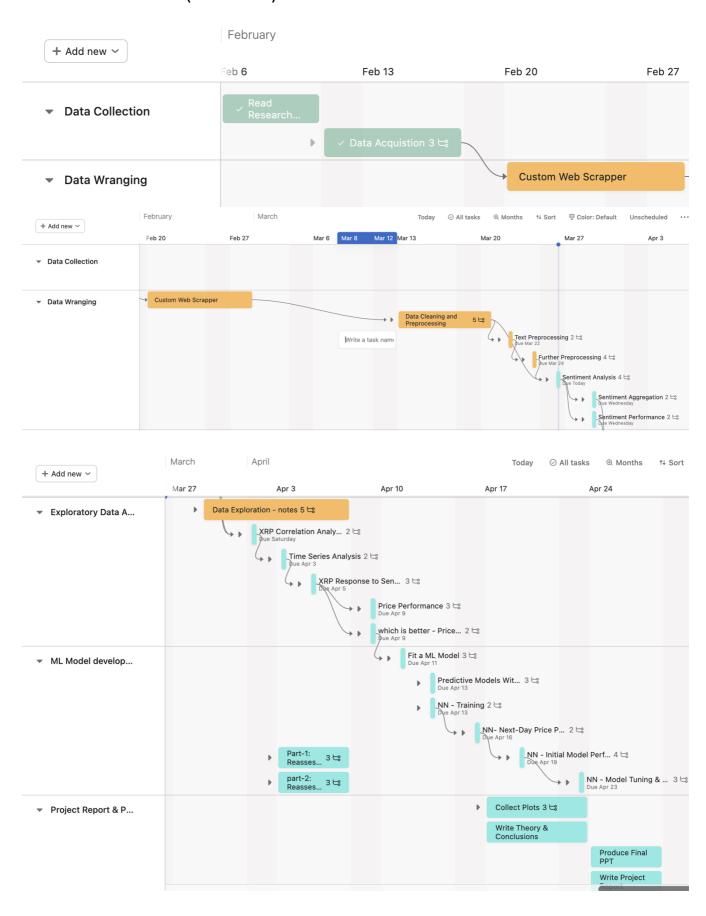
# Outline the project with the tasks and their status.

Tool : Asana

Cryptocurrency price prediction by Sentiment analysis ⌄ ☆ ○ Set status

Overview   List   Board   Timeline   Calendar   Dashboard   Messages   Files

Filter   Sort   Hide

## Data Collection
- Read Research paper — Complete — Feb 6 – 10
- Data Acquistion — Complete — Feb 11 – 17
- + Add task

## Data Wranging
- Custom Web Scrapper — In Progress — Yb — Feb 20 – 28
- Data Cleaning and Preprocess — In Progress — ch — Mar 13 – 20 — 5
- Text Preprocessing — In Progress — S — Mar 21 – 22 — 2
- Further Preprocessing — In Progress — dp — Mar 23 – 24 — 4
- Sentiment Analysis — To Do — Mar 25 – Today — 4
- Sentiment Performance — To Do — dp — Mar 27 – 28 — 2
- Sentiment Aggregation — To Do — Yb — Mar 27 – 28 — 2
- + Add task

## Exploratory Data Analysis
- Data Exploration - notes — In Progress — S — Mar 29 – Apr 7 — 5
- XRP Correlation Analysis — To Do — Mar 29 – 31 — 2
- Time Series Analysis — To Do — Yb — Apr 1 – 2 — 2
- XRP Response to Sentiment — To Do — S — Apr 3 – 4 — 3
- Price Performance — To Do — dp — Apr 5 – 7 — 3
- which is better - Price Range or Exact Price — To Do — ch — Apr 5 – 7 — 2
- + Add task

## ML Model development
- Part-1: Reassess Project Goals — To Do — Apr 3 – 7 — 3
- part-2: Reassess Project Goals — To Do — Yb — Apr 3 – 7 — 3
- Fit a ML Model — To Do — Apr 8 – 10 — 3
- Predictive Models With Deep Neural Nets — To Do — dp — Apr 11 – 13 — 3
- NN - Training — To Do — ch — Apr 11 – 13 — 2
- NN- Next-Day Price Prediction — To Do — Apr 14 – 16 — 2
- NN - Initial Model Performance — To Do — Yb — Apr 17 – 19 — 4
- NN - Model Tuning & Evaluation — To Do — Apr 20 – 23 — 3
- + Add task

## Project Report & Presentati... +
- Collect Plots — To Do — S — Apr 17 – 23 — 3
- Write Theory & Conclusions — To Do — Apr 17 – 23
- Write Project Report — To Do — dp — Apr 24 – 28
- Produce Final PPT — To Do — dp — Apr 24 – 28
- + Add task

+ Add section

# Section Wise TimeLine (Gantt Chart)

## February

| | Feb 6 | Feb 13 | Feb 20 | Feb 27 |
|---|---|---|---|---|

**▼ Data Collection**
- ✓ Read Research...
- ✓ Data Acquistion 3
- Custom Web Scrapper

**▼ Data Wranging**

---

| February | March | Today | ⊘ All tasks | ⊕ Months | ↑↓ Sort | ▽ Color: Default | Unscheduled | ⋯ |
|---|---|---|---|---|---|---|---|---|
| Feb 20 | Feb 27 | Mar 6 | Mar 8 Mar 12 Mar 13 | Mar 20 | Mar 27 | Apr 3 | | |

**▼ Data Collection**

**▼ Data Wranging**
- Custom Web Scrapper
- Write a task name
- Data Cleaning and Preprocessing 5
- Text Preprocessing 2 — Due Mar 22
- Further Preprocessing 4 — Due Mar 24
- Sentiment Analysis 4 — Due Today
- Sentiment Aggregation 2 — Due Wednesday
- Sentiment Performance 2 — Due Wednesday

---

| March | April | Today | ⊘ All tasks | ⊕ Months | ↑↓ Sort |
|---|---|---|---|---|---|
| Mar 27 | Apr 3 | Apr 10 | Apr 17 | Apr 24 | |

**▼ Exploratory Data A...**
- Data Exploration - notes 5
- XRP Correlation Analy... 2 — Due Saturday
- Time Series Analysis 2 — Due Apr 3
- XRP Response to Sen... 3 — Due Apr 5
- Price Performance 3 — Due Apr 9
- which is better - Price... 2 — Due Apr 9

**▼ ML Model develop...**
- Fit a ML Model 3 — Due Apr 11
- Predictive Models Wit... 3 — Due Apr 13
- NN - Training 2 — Due Apr 13
- NN- Next-Day Price P... 2 — Due Apr 16
- NN - Initial Model Perf... 4 — Due Apr 19
- NN - Model Tuning & ... 3 — Due Apr 23
- Part-1: Reasses... 3
- part-2: Reasses... 3

**▼ Project Report & P...**
- Collect Plots 3
- Write Theory & Conclusions
- Produce Final PPT
- Write Project Report

## List Of Task with Assignments followed by dependency:

| Sr no | Parent Task | Name | Assignee | Due Date |
|---|---|---|---|---|
| 1 | **Data Collection** | Read Research paper | Ashlesh Khajbage | 2023-02-10 |
| 2 | | Data  Acquistion | Ashlesh Khajbage | 2023-02-17 |
| 3 | **Data Wranging** | Custom Web Scrapper | Yeshwanth bolikonda | 2023-02-28 |
| 4 | | Data Cleaning and Preprocessing | chandanapoloju | 2023-03-20 |
| 5 | | Text Preprocessing | Swastika Pandit | 2023-03-22 |
| 6 | | Further Preprocessing | Deepti P | 2023-03-24 |
| 7 | | Sentiment Analysis | Ashlesh Khajbage | 2023-03-26 |
| 8 | | Sentiment Performance | Deepti P | 2023-03-28 |
| 9 | | Sentiment Aggregation | Yeshwanth bolikonda | 2023-03-28 |
| 10 | **Exploratory Data Analysis** | Data  Exploration - notes | Swastika Pandit | 2023-04-07 |
| 11 | | XRP Correlation Analysis | Ashlesh Khajbage | 2023-03-31 |
| 12 | | Time Series Analysis | Yeshwanth bolikonda | 2023-04-02 |
| 13 | | XRP Response to Sentiment | Swastika Pandit | 2023-04-04 |
| 14 | | Price Performance | Deepti P | 2023-04-07 |
| 15 | | which is better - Price Range or Exact | chandanapoloju | 2023-04-07 |
| 16 | **ML Model development** | Part-1: Reassess Project Goals | Ashlesh Khajbage | 2023-04-07 |
| 17 | | part-2: Reassess Project Goals | Yeshwanth bolikonda | 2023-04-07 |
| 18 | | Fit a ML Model | Ashlesh Khajbage | 2023-04-10 |
| 19 | | Predictive Models With Deep Neural Nets | Deepti P | 2023-04-13 |
| 20 | | NN - Training | chandanapoloju | 2023-04-13 |
| 21 | | NN- Next-Day Price Prediction | Ashlesh Khajbage | 2023-04-16 |
| 22 | | NN - Initial Model Performance | Yeshwanth bolikonda | 2023-04-19 |
| 23 | | NN - Model Tuning & Evaluation | Ashlesh Khajbage | 2023-04-23 |
| 24 | **Project Report & Presentation** | Collect Plots | Swastika Pandit | 2023-04-23 |
| 25 | | Write Theory & Conclusions | Ashlesh Khajbage | 2023-04-23 |
| 26 | | Write Project Report | Deepti P | 2023-04-28 |
| 27 | | Produce Final PPT | Deepti P | 2023-04-28 |

| Sr no | Blocked By (Dependencies) | Blocking (Dependencies) | Status |
|---|---|---|---|
| 1 | | | Complete |
| 2 | | Custom Web Scrapper | Complete |
| 3 | Data Acquistion | Data Cleaning and Preprocessing | In Progress |
| 4 | Custom Web Scrapper | Sentiment Analysis,Text Preprocessing | In Progress |
| 5 | Data Cleaning and Preprocessing | Further Preprocessing | In Progress |
| 6 | Text Preprocessing | Sentiment Analysis | In Progress |
| 7 | Data Cleaning and Preprocessing,Further Preprocessing | Sentiment Aggregation,Sentiment Performance | |
| 8 | Sentiment Analysis | XRP Correlation Analysis | |
| 9 | Sentiment Analysis | XRP Correlation Analysis | |
| 10 | | | In Progress |
| 11 | Sentiment Performance,Sentiment Aggregation | Time Series Analysis | To Do |
| 12 | XRP Correlation Analysis | XRP Response to Sentiment | To Do |
| 13 | Time Series Analysis | Price Performance,which is better - Price Range or Exact Price | To Do |
| 14 | XRP Response to Sentiment | | To Do |
| 15 | XRP Response to Sentiment | Fit a ML Model | To Do |
| 16 | | | To Do |
| 17 | | | To Do |
| 18 | which is better - Price Range or Exact Price | | To Do |
| 19 | | | |
| 20 | | NN- Next-Day Price Prediction | |
| 21 | NN - Training | NN - Initial Model Performance | To Do |
| 22 | NN- Next-Day Price Prediction | NN - Model Tuning & Evaluation | To Do |
| 23 | NN - Initial Model Performance | | To Do |
| 24 | | | To Do |
| 25 | | | To Do |
| 26 | | | To Do |
| 27 | | | To Do |